

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

GENES IN THE NON-RECOMBINING
REGION OF THE Y CHROMOSOME

GOVERNMENT SUPPORT

The invention described herein was made in whole or in
5 part with government support under Grant Number HG00257
awarded by the National Institutes of Health. The United
States Government has certain rights in the invention.

RELATED APPLICATIONS

This application claims the benefit of U.S.
10 Provisional Application No. 60/041,877, filed April 11,
1997, entitled "Genes in the Non-Recombining Region of the
Y Chromosome" by Bruce T. Lahn and David C. Page. The
entire teachings of the above referenced application is
expressly incorporated herein by reference.

15 BACKGROUND OF THE INVENTION

The human Y chromosome is distinguished from all other
nuclear chromosomes by four characteristics: the absence of
recombination, its presence in males only, its common
ancestry and persistent meiotic relationship with the X
20 chromosome, and the tendency of its genes to degenerate
during evolution (J. J. Bull, *Evolution of Sex Determining
Mechanisms* (Benjamin Cummings, Menlo Park, CA, 1983); J. A.
Graves, *Annu. Rev. Genet.* 30:233 (1996); B. Charlesworth,
Curr. Biol. 6:149 (1996); W. R. Rice, *BioScience*, 46, 331

-2-

(1996)). To be precise, these distinctive characteristics apply only to the non-recombining portion or region of the Y chromosome (NRY), which comprises 95% of the human Y chromosome. The remaining 5% of the chromosome is composed of two pseudoautosomal regions that maintain sequence identity with the X chromosome by meiotic recombination (H. J. Cooke et al., *Nature* 317:687 (1985); M. C. Simmler et al., *Nature* 317:692 (1985); D. Freije et al., *Science* 258:1784 (1992); G. A. Rappold, *Hum. Genet.* 92:315 (1993)).

Given the NRY's peculiar characteristics, one might expect its gene content to be idiosyncratic. Since discovery of the Y chromosome in 1923, its gene content has been the subject of speculation. By the middle of this century, while studies of human pedigrees had identified many traits exhibiting autosomal or X-linked inheritance, no convincing cases of Y-linked inheritance could be found (T. S. Painter, *J. Exp. Zool.* (1923); C. Stern, *Am. J. Hum. Genet.* 9:147 (1957)). As a result, consensus began to emerge that the Y chromosome carried few, if any, genes. In 1959, reports of XO females and XXY males established the existence of a sex-determining gene on the human Y chromosome (P. A. Jacobs et al. *Nature* 183:302 (1959); C. E. Ford et al., *Lancet*, i:711 (1959)), but this was perceived as a special case on a generally desolate chromosome. Opinions began to change only during the past decade, when eight NRY transcription units (or families of closely related transcription units) were identified, most during regionally focused, positional cloning experiments (D. C. Page et al., *Cell* 51:1091 (1987); A. H. Sinclair et al., *Nature* 346:240-244 (1990); J. Arnemann et al., *Genomics* 11: 108 (1991); E. C. Salido et al., *Am. J. Hum. Genet.* 50:303 (1992); E. M. Fisher et al., *Cell* 63:1205 (1990); K. Ma et al., *Cell* 75:1287 (1993); A. I. Agulnik et al., *Hum. Mol. Genet.* 3:879 (1994); R. Reijo et al., *Nat.*

-3-

Genet. 10:383 (1995)). It was not known if there were more genes in the NRY.

SUMMARY OF THE INVENTION

A systematic search of the non-recombining region of the human Y chromosome (NRY) has identified 12 novel genes or gene families. All 12 novel genes, and six of eight NRY genes or families previously isolated by less systematic means, fall into two classes. The first class of genes exists in one copy and is expressed in many organs; they have functional X homologs that escape X inactivation, as predicted for genes involved in Turner (XO) syndrome. The second class consists of Y-chromosomal gene families expressed specifically in testes, and may account for infertility among men with Y deletions.

The genes described herein, portions of the genes and DNA which hybridizes to genes or gene portions described are useful in diagnostic methods, such as a method to identify individuals in whom all or a portion of a gene or genes of the NRY is missing or altered. For example, Y chromosomal DNA from males with a known condition, such as infertility or reduced sperm count, can be assessed, using the gene(s) described herein, or characteristic portions thereof, to determine whether their DNA lacks some or all of the gene(s) described herein or contains an altered gene(s) (e.g., a gene in which there is a deletion, substitution, addition or mutation, compared to the sequences presented herein). Y chromosomal DNA (e.g., from a male with reduced sperm count or viability) can be assessed, using DNA described herein or DNA which hybridizes to DNA described herein, to determine whether the condition is associated with or caused by the occurrence of the gene or the gene alteration. For example, the presence or absence of all or a portion of a gene or genes shown to be necessary for fertility or

-4-

adequate sperm count can be assessed, using DNA which hybridizes to the gene or genes of interest to determine the basis for their infertility or reduced sperm count. In one embodiment, the occurrence of one or more Y-specific genes or a characteristic portion of one or more Y-specific genes is assessed in Y chromosomal DNA. In another embodiment, deletion or alteration of one of the testis-specific (Y-specific) genes described is assessed, such as by a hybridization method in which DNA which hybridizes to one of the Y-specific genes described herein or a characteristic portion thereof is used to assess a DNA sample obtained from a male who has a reduced sperm count. Lack of hybridization of the Y-specific DNA used to DNA in the sample indicates that the gene is not present in sample DNA or is present in an altered form which does not hybridize to Y-specific DNA of the present invention. In another embodiment, an X-homologous gene or genes present on the NRY can be used to determine whether the gene is present in an individual or if it occurs in an altered form in the individual. Using known methods, such as hybridization methods, X or Y chromosomal DNA from an individual can be assessed for the presence or absence of one or more of the X-homologous genes or a characteristic portion of one or more X-homologous genes. X or Y chromosomal DNA can also be assessed for the presence or absence of an altered form of one or more of the X-homologous genes described. In the present methods, DNA can be analyzed for the occurrence of Y-specific DNA, X-homologous genes or both. For example, a "battery" or group of DNA probes (sequences) can be used to analyze sample DNA; the probes can include Y-specific DNA probes (e.g., DNA which hybridizes to a Y-specific gene), X-homologous gene probes (e.g., DNA which hybridizes to an X-homologous gene) or both types of probes. DNA described herein is also useful as primers in an amplification

-5-

method, such as PCR, useful for identifying and amplifying Y-specific DNA or X-homologous genes in a sample (e.g., Y chromosomal DNA). Further, proteins or peptides encoded by the DNA described herein, such as proteins or peptides encoded by an X-homologous gene or proteins or peptides encoded by testis-specific DNA (a testis-specific gene), can be assessed in samples. This can be carried out, for example, using antibodies which recognize proteins or peptides of the present invention (proteins or peptides encoded by DNA described herein).

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a gene map of the non-recombining region of the Y chromosome.

Figure 2 shows the amino acid sequence alignments of the chromodomain (SEQ ID NO.: 1-6) and putative catalytic domain (SEQ ID NO.: 7-12) of human CDY genes with their respective homologs. Amino acid identities are indicated by black shading and for each protein, the first and last amino acid residues are numbered (with respect to the initiator methionine) and the total length of the protein is indicated. Chromodomain: SEQ ID NO.: 1, CDY (human); SEQ ID NO.: 2, HP1 (Drosophila); SEQ ID NO.: 3, Polycomb (Drosophila); SEQ ID NO.: 4, CHD1 (Drosophila); SEQ ID NO.: 5, Su(var) 3-9 (Drosophila); SEQ ID NO.: 6, PDD1 (Tetrahymena); SEQ ID NO.: 7; Covalent modification domain: SEQ ID NO.: 8, CDY (human); SEQ ID NO.: 9, Enoyl-CoA Hydratase (Human); SEQ ID NO.: 10, 4-CBA-CoA dehalogenase (Arthrobacter); SEQ ID NO.: 11, Crotonase (C. acetobutylicum); SEQ ID NO.: 12, Naphthoate synthase (E. coli).

Figures 3A and 3B are the nucleic acid sequence of DBX (long and short transcripts, SEQ ID NO: 13 and SEQ ID NO: 14, respectively) and the encoded amino acid sequences (SEQ ID NO: 15 and SEQ ID NO.: 16, respectively), DBY (SEQ ID

-6-

NO: 17) and the encoded amino acid sequence (SEQ ID NO: 18). Dots in the DBX DNA and protein sequences indicate that the nucleic acids or amino acid residues are the same as those represented for DBY; dashes indicate a missing
5 nucleic acid or amino acid residue.

Figures 4A and 4B present the nucleic acid sequences for three forms of TPRY (short, medium and long, SEQ ID NO: 19, SEQ ID NO: 20 and SEQ ID NO: 21, respectively) and the encoded amino acid sequences for the short, medium and long
10 forms (SEQ ID NO: 22, SEQ ID NO.: 23 and SEQ ID NO: 24, respectively).

Figure 5 presents the nucleic acid sequences of TB4X (SEQ ID NO: 25) and TB4Y (SEQ ID NO: 26) and the encoded amino acid sequences (SEQ ID NO: 27 and SEQ ID NO: 28,
15 respectively). Dots in the TB4X DNA and protein sequences indicate that the nucleic acids or amino acid residues are the same as those represented for TB4Y.

Figure 6 represents the nucleic acid sequences of EIF1AX (SEQ ID NO: 29) and EIF1AY (SEQ ID NO: 30) and the
20 encoded amino acid sequences (SEQ ID NO: 31 and SEQ ID NO: 32, respectively).

Figures 7A - 7D represent the nucleic acid sequences of DFFRX (SEQ ID NO: 33) and DFFRY (SEQ ID NO: 34) and the encoded amino acid sequences (SEQ ID NO: 35 and SEQ ID NO:
25 36, respectively).

Figure 8 represents the nucleic acid sequences of CDYa (SEQ ID NO: 37) and CDYb (SEQ ID NO: 38) and the encoded amino acid sequences (SEQ ID NO: 39 and SEQ ID NO: 40,
respectively).

30 Figure 9 represents the nucleic acid sequences of BPY1 (SEQ ID NO: 41) and the encoded amino acid sequence (SEQ ID NO: 42).

Figure 10 represents the nucleic acid sequence of BPY2 (SEQ ID NO: 43) and the encoded amino acid sequence (SEQ ID
35 NO: 44).

-7-

Figure 11 represents the nucleic acid sequences of XKRY (SEQ ID NO: 45) and the encoded amino acid sequence (SEQ ID NO: 46).

Figure 12 represents the nucleic acid sequences of PTPRY (SEQ ID NO: 47) and the encoded amino acid sequence (SEQ ID NO: 48).

Figure 13 is the nucleic acid sequence of TTY1 (SEQ ID NO: 49).

Figure 14 is the nucleic acid sequence of TTY2 (SEQ ID NO: 50).

Figure 15 shows the nucleic acid sequence of the human CDY Like (CDYL) gene, which is the human autosomal homolog of CDY, located on chromosome 6p and expressed ubiquitously.

Figure 16 shows the nucleic acid sequence of the mouse Cdyl (CDY like) gene, which is the mouse ortholog of human CDYL, located on chromosome 13 and expressed predominantly in the testis. A longer transcript of the gene is ubiquitously expressed.

Figures 17A - 17C show the nucleic acid sequences of human Variably Charged Protein family members VCP2r, VCP8r and VCP10r, which are expressed in the testis and highly polymorphic.

Figure 17A is the nucleic acid sequence of VCP2r.

Figure 17B is the nucleic acid sequence of VCP8r.

Figure 17C is the nucleic acid sequence of VCP10r.

DETAILED DESCRIPTION OF THE INVENTION

Y chromosome genes, classed as genes having X homologues and testis-specific (Y-specific) genes, are the subject of the invention described herein, as are DNA which hybridize to (are complementary to) all or characteristic portions of the Y chromosome genes, the encoded products (e.g., proteins, peptides, glycoproteins), antibodies and methods of diagnosis or treatment in which the genes,

-8-

complementary DNA, encoded proteins or antibodies are used. As described herein, fragments that hybridized to Y chromosomal DNA were selected and then their nucleotide sequences determined. It was expected that these sequence
5 fragments would represent a redundant sampling of a much smaller set of genes. Computer analysis revealed that 577 fragments corresponded to known Y genes, including seven of eight NRY genes and all eight pseudoautosomal genes previously reported. These findings suggested that the
10 2539 sequence fragments represented the great majority of all Y-chromosomal genes. After further analysis, both to eliminate human repetitive sequences and to assemble overlapping fragments into contigs, 912 novel and non-overlapping sequences were hybridized to Southern blots
15 of human genomic DNAs. 308 sequences that detected at least one prominent male-specific fragment were judged likely to derive from the NRY, and for each work was carried out to isolate cDNA clones from a human testis library, as described in Example 1. Nucleotide sequencing
20 of cDNA clones, and rescreening of libraries as necessary, yielded full-length cDNA sequences for ten novel NRY genes or families, and partial cDNA sequences for two additional ones (Table and Figures 1 - 14).

TABLE: 12 Novel Genes or Families in the NRY

Gene Symbol	Gene Name	Tissue Expression	Multi-copy on Y	X homolog	Escape x Inactivation
DBY	Dead Box Y	ubiquitous		DBX	yes
TB4Y	Thymosin β 4, Y isoform	ubiquitous		TB4X	yes
EIF1AY	Translation Initiation Factor 1A, Y isoform	ubiquitous		EIF1AX	yes
TPRY	TPR motif Y	ubiquitous		TPRX	yes
DDFRY	Drosophila Fat Facets Related Y	ubiquitous		DDFRX	yes
CDY	Chromodomain Y	testis	yes		
BPY1	Basic Protein Y 1	testis	yes		
BPY2	Basic Protein Y 2	testis	yes		
XKRY	XK Related Y	testis	yes		
PTPRY	Protein-Tyrosine Phosphatase Related Y	testis	yes		
TTY1	Testis Transcript Y 1	testis	yes		
TTY2	Testis Transcript Y 2	testis	yes		

-10-

All 12 novel genes were localized on the Y chromosome, as described in Example 2. Figure 1 is a gene map of NRY. As shown, the Y chromosome consists of a large non-recombining region (NRY; euchromatin plus heterochromatin) flanked by pseudoautosomal regions (pter, short arm telomere; qter, long arm telomere). The NRY is divided into 43 ordered intervals (1A1A through 7) which are defined by naturally occurring deletions (D. Vollrath, et al., *Science* 258:52 (1992)). Listed immediately above the Y chromosome in Figure 1 are nine NRY genes with functional X homologs; novel genes are boxed. Indicated immediately below the Y chromosome are 11 testis-specific genes or families, some with multiple locations. It is likely that some testis-specific families have members in additional deletion intervals; the locations indicated are representative, but are not necessarily exhaustive. At the bottom of Figure 1 are shown NRY regions implicated, by deletion mapping, in sex determination, germ cell tumorigenesis (gonadoblastoma), stature, and spermatogenic failure (K. Ma et al., *Cell* 75:1287 (1993); R. Reijo et al., *Nat. Genet.* 10:383 (1995); P. H. Vogt et al., *Hum. Mol. Genet.* 5:933 (1996); J. L. Pryor et al., *New England J. Med.* 336:534 (1997); K. Tsuchiya et al., *Am. J. Hum. Genet.* 57:1400 (1995); P. Salo et al., *Hum. Genet.* 95:283 (1995)). Euchromatic regions that are made up, at least partially, of Y-specific repeats are drawn in grey. *AMELY*, which appears to fall within such a repeat-containing region, is actually located in a sub-region of 4A that is not repetitive.

Expression of the 12 novel genes was assessed in diverse human tissues, by Northern blotting. Autoradiograms were produced by hybridizing ³²P-labeled cDNA probes to Northern blots of poly(A)⁺ RNAs (2 µg/lane) from human tissues (Clontech, Palo Alto, CA). Probes employed were cDNA clones, full-length (most genes) or

-11-

partial (*DBY*, nucleotides 1476-2319 of GenBank AF000985; *TPRY*, nucleotides 861-1768 of GenBank AF000996; *DDFRY*, nucleotides 8604-9878 of GenBank AF000986). Blots were hybridized at 65°C in Church's buffer (0.5 M Na₂PO₄ at pH7.5, with 7% SDS), and washed at 65°C in 1X SSC and 0.1% SDS. *DBY*, *TB4Y*, *EIF1AY* and *DDFRY* probes cross-hybridize to transcripts derived from their X homologs. For all five X-homologous genes (*DBY*, *TPRY*, *TB4Y*, *EIF1AY* and *DDFRY*), expression was tested and confirmed in three male tissues (brain, prostate and testis) by RT-PCR using Y-specific primers.

The novel genes encode an assortment of proteins and are dispersed throughout the euchromatic portions of the NRY. Nonetheless, all 12 genes fall into two discrete classes: 1) X-homologous genes and 2) testis-specific, Y-specific gene families (Table).

The X-homologous genes share the following characteristics: each has a homolog on the X chromosome encoding an extremely similar but nonidentical protein isoform, each is expressed in a wide range of human tissues (is not testis-specific), and each appears to exist in a single copy on the NRY. There are five novel representatives of this X-homologous class:

1. *DBY* encodes a novel "DEAD box" protein, perhaps an RNA helicase involved in translation initiation (P. Linder, et al., *Nature*, 337, 121 (1989); R.-Y. Chuang, P. L. Weaver, Z. Liu, T.-H. Chang, *Science*, 275, 1468 (1997)). The *DBY* protein is 91% identical to *DBX*, encoded by a homologous gene on the human X chromosome.
2. *TPRY* encodes a novel protein containing 10 tandem "TPR" motifs, a protein-protein interaction domain found in the products of the yeast *SSN6/CYC8*, *CDC16*, and *CDC23* genes, among others (R. S. Sikorski, M. S. Boguski, M. Goebel, P. Hieter, *Cell*, 60, 307 (1990); D. Tzamarias, K. Struhl, *Genes Dev*, 9, 821 (1995)). Differential splicing may

-12-

generate TPRY isoforms that differ at their carboxy termini. The amino terminal portion of the TPRY protein is 83% identical to TPRX, encoded by an homologous gene on the X chromosome.

5 3. *TB4Y* encodes a 44 amino acid protein that differs at only three residues from thymosin β_4 , which functions in actin sequestration (H. Gondo, et al., *J. Immunol.* 139:3840 (1987); D. Safer, M. Elzinga, V. T. Nachmias, *J Biol Chem*, 266, 4029 (1991)), and we found is located on the X. It is
10 proposed that the X-linked gene encoding thymosin β_4 be called *TB4X*.

4. *EIF1AY* encodes a Y-linked isoform of translation initiation factor 1A (eIF-1A) (T. E. Dever, et al., *J Biol Chem*, 269, 3212 (1994); J. W. Hershey, *Annu. Rev. Biochem.*
15 60, 717 (1991)), which we discovered is located on the X. It is proposed that the X-linked gene encoding eIF-1A be called *EIF1AX*. The amino acid sequences of the X and Y-encoded proteins are 97% identical.

5. *DFFRY* encodes a Y-linked isoform of *DFFRX*, a recently
20 described X-linked protein. A Y-linked homolog was detected previously, but had been thought to be a pseudogene. The human *DFFRX* and *DFFRY* proteins, which are 91% identical, are homologous to the *Drosophila fat-facets* gene product, a deubiquinating enzyme required for eye
25 development and oogenesis (M. H. Jones, et al., *Hum Mol Genet* 5, 1695 (1996); J. A. Fischer-Vize, G. M. Rubin, R. Lehmann, *Development*, 116, 985 (1992); Y. Huang, R. T. Baker, J. A. Fischer-Vize, *Science*, 270, 1828 (1995)).

The second group of novel NRY genes, the testis-specific, Y-specific gene families, share a very different
30 set of characteristics: each appears to be expressed specifically in testes and each appears to exist in multiple copies on the NRY, as judged by i) the number and intensity of hybridizing fragments on genomic Southern
35 blots or ii) multiple map locations on the Y. We report

-13-

five novel testis-specific, Y-specific gene families with full-length cDNA sequences:

1. The *CDY* family encodes proteins with an amino-terminal "chromodomain," a chromatin binding motif (T. C. James, S. C. Elgin, *Mol Cell Biol*, 6, 3862 (1986); B. Tschiersch, et al., *EMBO J*, 13, 3822 (1994); R. Paro, D. S. Hogness, *Proc Natl Acad Sci U S A*, 88, 263 (1991); D. G. Stokes, K. D. Tartof, R. P. Perry, *Proc Natl Acad Sci U S A*, 93, 7137 (1996); M. T. Madireddi, et al., *Cell*, 87, 75 (1996)) (Figure 3). The carboxy-terminal half shows striking amino acid similarity, over a region of more than 200 residues, to nearly the full length of several enzymes, both prokaryotic and eukaryotic (M. Kanazawa, et al., *Enzyme Protein*, 47, 9 (1993); A. Schmitz, K. H. Gartemann, J. Fiedler, E. Grund, R. Eichenlaub, *Appl. Environ. Microbiol.* 258, 4068 (1992); Z. L. Boynton, G. N. Bennet, F. B. Rudolph, *J Bacteriol*, 178, 3015 (1996); V. Sharma, K. Suvarna, R. Meganathan, M. E. Hudspeth, *J Bacteriol*, 174, 5057 (1992); P. M. Palosaari, et al., *J Biol Chem*, 266, 10750 (1991)). The reactions catalyzed by these homologs are diverse, but in each case the substrate contains cofactor A (CoA) attached to a carbonyl group, and an alkoxide intermediate is formed. The unprecedented combination of a chromodomain and a putative CoA-substrate enzyme in a single polypeptide suggests that, in vivo, *CDY* proteins may catalyze covalent modification of DNA or chromosomal proteins, perhaps during spermatogenesis.
2. The *BPY1* genes encode a basic protein, 125 residues long, with little sequence similarity to known proteins. The encoded protein is rich in serine, lysine, arginine, and proline and has a pI of 9.4. Southern blotting studies revealed homologous sequences on the human X chromosome, but screening of cDNA libraries has failed to yield X-derived clones.

-14-

3. The *BPY2* genes encode a second basic protein, 106 residues in length, without obvious sequence similarity to *BPY1* or other known proteins. The pI of *BPY2* is 10.0.

4. The *XKRY* genes encode a protein with sequence
5 similarity to *XK*, a putative membrane transport protein defective in McLeod syndrome (M. Ho, et al., *Cell*, 77, 869 (1994)).

5. The *PTPRY* genes encode a protein with weak homology to a putative protein-tyrosine phosphatase (PTPase) in the
10 mouse (W. Hendriks, et al., *J Cell Biochem*, 59, 418 (1995)). Two additional families of testis-specific transcription units, referred to as *TTY1* and *TTY2*, have been identified. The sequences represented in Figures 14 and 15 are being assessed for open reading frames.

15 It appears that conventional single-copy genes, commonplace elsewhere in the genome, are quite uncommon in the NRY. Indeed, the two classes of NRY genes suggested by the systematic search described herein accommodate not only the 12 genes reported here, but also six of eight
20 previously identified NRY genes. *SRY*, a Y-specific gene that triggers the male pathway of sexual differentiation, is expressed in testes, and exists in only one copy in the NRY. *AMELY*, which has an X-linked homolog *AMELX*, is expressed only in the developing tooth bud. The X
25 inactivation status of *AMELX* is unknown.

Also described herein are five additional genes and their sequences (Figures 15, 16, 17A - 17C): human *CDY* Like (*CDYL*), which is the human homolog of *CDY*; it is on chromosome 6p and expressed ubiquitously; mouse *Cdyl* (*CDY*
30 like), which is the mouse ortholog of human *CDYL*; it is on chromosome 13 and expressed predominantly in testis and also has a longer transcript that is expressed ubiquitously; and human *VCP* (Variably Charged Protein) family, which is a family of genes on the X chromosome that
35 are homologous to *BPY1*, expressed in the testis and highly

-15-

polymorphic. Human CDY, human CDYL and mouse Cdyl have been shown to be histone acetyltransferases by *in vitro* assays. Human CDY is a candidate for the Azoospermia Factor (AZF) because it is within the AZFc region that is commonly deleted in infertile men. Chemicals that block the enzymatic activity of any of these genes are candidate male contraceptives.

Inhibitors of the enzymatic activity of these genes, such as the human CDY gene, can be identified through an *in vitro* assay. For example, the protein encoded by one of the genes (e.g., CDY-encoded protein) can be produced, such as by recombinant means (e.g., in bacterial cells containing a vector or plasmid which includes the gene to be expressed), and obtained. The effect of a candidate inhibitor (drug) on the enzymatic activity of the protein can be assessed by combining the candidate inhibitor with the protein, a substrate of its enzymatic activity (e.g., histones) acetyl CoA (e.g., radiolabelled acetyl CoA) and other assay components (e.g., an appropriate physiological solution or buffer), to produce a combination. The combination is maintained under conditions under which the enzymatic activity of the protein is maintained and appropriate for the protein to act upon/interact with its substrate (e.g., for the CDY gene to retain its histone acetyltransferase activity). As a result, the substrate is acted upon by the protein if the candidate inhibitor does not inhibit the protein and the protein acts upon the substrate. If the substrate is not acted upon by the protein, this is an indication that the candidate inhibitor is an inhibitor of the protein. For example, if a histone acetyltransferase, such as CDY-encoded protein is inhibited by a candidate inhibitor, its histone acetyltransferase activity will be blocked. If radiolabelled acetyl CoA is used, transfer of the radiolabelled acetyl group to the enzyme substrate (histones) is inhibited (will not occur or

-16-

will occur to a lesser extent than occurs in the absence of the candidate inhibitor). Whether transfer occurs can be assessed by determining the location of radiolabelled acetyl groups from acetyl CoA. If the histone substrates
5 are not radiolabelled or are radiolabelled to a lesser extent in the presence of a candidate inhibitor (than in its absence), the candidate inhibitor is an inhibitor of the protein. Inhibitors identified in this way can be further assessed in additional *in vitro* assays or in *in vivo* assays (e.g., in an appropriate animal model).
10

To interpret the observation that these X-homologous and multi-copy, testis-specific groups account for 18 of 20 known NRY genes or families, we postulate that the NRY's evolution was dominated by two strategies. The first
15 strategy favors conservation of certain existing genes and the second favors the acquisition of a class of novel genes: 1) The X-homologous genes probably reflect the common ancestry of the X and Y chromosomes, and selective pressures to maintain comparable expression of genes in
20 males and females. 2) The abundance of testis-specific families may have resulted from the NRY's selectively retaining and amplifying genes that enhance male reproductive fitness.

1) Dosage compensation and X-Y homology. Experts
25 agree that the mammalian X and Y chromosomes evolved from autosomes, with nearly all ancestral gene functions deteriorating on the non-recombining portion of the emerging Y chromosome while being maintained on the nascent X chromosome (J. J. Bull, *Evolution of Sex Determining*
30 *Mechanisms* (Benjamin Cummings, Menlo Park, CA, 1983); J. A. Graves, *Annu. Rev. Genet.* 30:233 (1996); B. Charlesworth, *Curr. Biol.* 6:149 (1996); W. R. Rice, *BioScience* 46:331 (1996)). Functional degeneration of the NRY would result in females having two, but males only one, copy of many
35 genes, creating the need for a mechanism to equalize

-17-

X-linked gene expression in the sexes. In mammals, a predominant solution to this problem is provided by X inactivation, the transcriptional silencing of one X chromosome in females.

5 However, the findings on X-homologous NRY genes described herein, combined with previous studies, illustrate the importance in human evolution of an alternative solution: preservation of homologous genes on both the NRY and the X chromosome, with both male and
10 female cells expressing two copies of such genes. A critical prediction of this model is that, in female cells, the X homologs should escape X inactivation. This is the case for all widely expressed X-linked genes with known NRY homologs, including the X homologs of five novel NRY genes
15 reported here (E. M. Fisher, et al., *Cell* 63:1205 (1990); A. I. Agulnik et al., *Hum. Mol. Genet.* 3:879 (1994); M. H. Jones et al., *Hum. Mol. Genet.* 5:1695 (1996); J. A. Fischer-Vize et al., *Development* 116:985 (1992); Y. Huang et al., *Science* 270:1828 (1995); A. Schneider-Gädicke et
20 al., *Cell* 57:1247 (1989)). A second prediction of this model is that the human X and Y encoded proteins should be functionally interchangeable even though the nucleotide sequences of their corresponding genes are considerably diverged. Indeed, each of the eight known X-NRY gene pairs
25 encode closely related isoforms, with 83 to 97% amino acid identity throughout their lengths; functional interchangeability has been demonstrated in the one case tested to date (M. Watanabe et al., *Nat. Genet.* 4:268
(1993)).

30 Turner syndrome is classically associated with an XO sex chromosome constitution. In 1965, Ferguson-Smith postulated that the Turner phenotype might be due to inadequate expression of X-Y common genes that escape X inactivation (M. A. Ferguson-Smith, *J. Med. Genet.* 2:142
35 (1965)). These "Turner genes" have yet to be identified

-18-

with certainty. However, there now exists a substantial collection of X-homologous NRY genes (Figure 1) which can be assessed for genes which contribute to or are responsible for the Turner phenotype. The potential role of *RPS4Y* and *RPS4X* in Turner syndrome is controversial (E. M. Fisher et al., *Cell* 63:1205 (1990); W. Just et al., *Hum. Genet.* 89:240 (1992)). At least one Turner gene maps to the Xp-Yp pseudoautosomal region (T. Ogata et al., *J. Med. Genet.* 30:918 (1993)). Seven of the eight known X-NRY gene pairs appear to be ubiquitously expressed, and at least three encode housekeeping proteins: an essential ribosomal protein (*RPS4*), an essential translation initiation factor (*eIF-1A*), and a modulator of actin polymerization (thymosin β 4). Perhaps some features of the XO phenotype (e.g., poor fetal viability) reflect inadequate expression of such housekeeping functions.

2) Male fitness and Y-specific, testis-specific genes. As first appreciated by R.A. Fisher, animal genomes may contain genes or alleles that enhance male reproductive fitness but are inconsequential or detrimental with respect to female fitness (R. A. Fisher, *Biol. Rev.* 6:345 (1931)). As Fisher recognized, selective pressures would tend to favor the accumulation of such genes in male-specific regions of genomes. Of course, male reproductive fitness depends critically on sperm production, the central task of the adult testis. Since the NRY is the only male-specific portion of the mammalian genome, it should have a unique tendency to accumulate male-benefit genes during evolution.

These principles are illustrated by several gene families on the human NRY. *De novo* deletions of the *DAZ* gene cluster on the human Y chromosome are associated with severe spermatogenic defects (R. Reijo et al., *Nat. Genet.* 10:383 (1995)), and in *Drosophila* the *DAZ* homolog *boule* is required for spermatogenesis (C. G. Eberhart et al., *Nature* 381:783 (1996)). The *DAZ* gene cluster on the human Y

-19-

chromosome arose, during primate evolution, by transposition and amplification of an autosomal gene. Likewise, two other testis-specific NRY gene families —YRRM and TSPY — may also be the result of the Y chromosome's having acquired and amplified autosomal genes (R. Saxena et al., *Nat. Genet.* 14:292 (1996); M. L. Delbridge et al., *Nat. Genet.* 15:131 (1997)). It is possible that the selective advantage conferred by the NRY's retaining and amplifying male fertility factors (from throughout the genome) accounts for the multitude of testis-specific gene families there. This may have been the preeminent force in shaping the NRY's gene repertoire, as it appears that the great majority of NRY transcription units are members of such testis-specific families. In the NRY, each of the testis-specific gene families has multiple members, 20 to 40 copies in the case of TSPY (E. Manz et al., *Genomics* 17: 726 (1993)), and perhaps as many as 20 copies in the case of YRRM (K. Ma et al., *Cell* 75:1287 (1993)). All together, the various Y-specific gene families may include as many as several hundred genes or copies. Though it is not known how many of these are functional, it seems likely that Y-specific, testis-specific gene families comprise the great majority of NRY transcription units.

Recent genetic studies underscore the importance of the human Y chromosome in fertility. Many men with spermatogenic failure, but who are otherwise healthy, have deletions of portions of the NRY (K. Ma et al., *Cell* 75: 1287 (1993); R. Reijo et al., *Nat. Genet.* 10:383 (1995); P. H. Vogt et al., *Hum. Mol. Genet.* 5:933 (1996); J. L. Pryor et al., *New England J. Med.* 336:534 (1997)). These findings suggested the existence of NRY genes that play critical roles in male germ cell development but are not required elsewhere in the body. Previous deletion mapping studies have implicated four regions of the NRY in either

-20-

spermatogenic failure or germ cell tumorigenesis, and in each of these four regions we now report novel candidate genes expressed specifically, or most abundantly, in testes (Figure 1). As shown in Figure 1, the region implicated in gonadoblastoma, stature and spermatogenic failure all contain novel candidate genes. Two of the three regions implicated in spermatogenic failure each contain one or more novel testis-specific genes. The third region implicated in spermatogenic failure (intervals 5B-5D) contains two X-homologous genes, *DBY* and *EIF1AY*, with abundant, testis-specific transcripts in addition to higher-molecular-weight, ubiquitous transcripts.

While X-homologous and testis-specific genes are somewhat intermingled within the NRY, clustering is evident (Figure 1). The geographic distribution of the two classes correlates quite well with previously identified sequence domains within the euchromatic NRY (D. Vollrath et al., *Science* 258:52 (1992); S. Foote et al., *Science* 258:60 (1992)). Ten of the 11 known testis-specific families map to previously identified regions of Y-specific repetitive sequences. The only exception is *BPY1*, which cross-hybridizes to the X chromosome and maps to a previously recognized region of X homology. Indeed, one or more testis-specific gene families are found in nearly all known regions of euchromatic Y repeats (Figure 1). Ironically, it had been widely assumed that these regions consisted of "junk" DNA, partly on theoretical grounds (B. Charlesworth, *Science* 251:1030 (1991); E. Seboun et al., *Cold Spring Harb. Symp. Quant. Biol.* 1:237 (1986)). To the contrary, the results presented here argue that these Y-specific repetitive regions contain the great majority of the NRY's transcription units (The only exception is *BPY1*, which cross-hybridizes to the X chromosome and maps to a previously recognized region of X homology). These regions may be the result of rampant gene amplification during

-21-

mammalian evolution. By contrast, none of the eight X-homologous genes map to the Y-repeat regions; all eight map to regions previously identified as consisting largely of single-copy (or in some cases X-homologous) sequences.

5 It is possible that, early in mammalian evolution, these regions of the NRY shared extensive sequence identity with the nascent X chromosome. The stage is now set for systematic evolutionary, biochemical and cell biological studies of the NRY, an idiosyncratic segment of the human

10 genome.

The present invention relates to isolated DNA and genes, present on (which occur on) the Y chromosome, whose sequences are provided herein, as well as characteristic portions of the DNA. It relates to additional nucleic

15 acid/nucleotide sequences which are not identical to the sequences presented herein but include substitutions or differences; DNA which includes substitutions or differences and encodes the same amino acid sequence as a DNA whose sequence is provided herein or includes

20 substitutions which do not alter the ability of a DNA probe or primer which hybridizes to DNA whose sequence is presented herein to hybridize to the DNA containing the substitutions or differences. It further relates to DNA which encodes a protein or peptide whose sequence is

25 presented herein. The present invention also includes the complements of the DNA sequences presented herein, DNA which hybridizes under stringent (high stringency) conditions to the DNA whose sequences are presented and to RNA transcripts. The invention further relates to encoded

30 proteins, peptides and other products (e.g., glycoproteins) and antibodies which are raised against or bind to proteins or peptides whose amino acid sequences are presented herein or are encoded by DNA whose sequences are provided. As used herein, the term isolated DNA which occurs on the non-

35 recombining region of the human Y chromosome refers to DNA

-22-

which has been obtained or removed from the human Y chromosome or DNA, produced by any means (e.g., recombinant techniques, synthetic methods), which has the sequence of such Y chromosome DNA. For example, isolated testis-specific DNA or isolated testis-specific DNA which occurs on the non-recombining region of the human Y chromosome is DNA which has been obtained or removed from the non-recombining region of the human Y chromosome or which has the sequence of such DNA and has been obtained or produced by any means.

Thus, this invention has application to several areas. It may be used diagnostically to identify males with reduced sperm count in whom a gene has been deleted or altered. It may also be used therapeutically in gene therapy treatments to remedy fertility disorders associated with deletion or alteration of a gene described. In one embodiment of a gene therapy method, a gene described herein, or a gene portion which encodes a functional protein, is introduced into a man whose sperm count is reduced and in whom the gene is expressed and the encoded protein replaces the protein normally produced or enhances the quantity produced. The present invention may also be useful in designing or identifying agents which function as a male contraceptive by inducing reduced sperm count. This invention also has application as a research tool, as the nucleotide sequences described herein have been localized to regions of the Y chromosome.

The present invention includes nucleotide sequences described herein, and their complements, which are useful as hybridization probes or primers for an amplification method, such as polymerase chain reaction (PCR), to show the presence, absence or disruption of the gene of the present invention. Probes and primers can have all or a portion of the nucleotide sequence (nucleic acid sequence) of a gene described herein or all or a portion of its

-23-

complement. For example, sequences shown in the Figures or Example 2 (SEQ ID NOS.: 1-84), as well as the complements thereof, can be used. The probes and primers can be any length, provided that they are of sufficient length and appropriate composition (appropriate nucleotide sequence) to hybridize to all or an identifying or characteristic portion of the gene described or to a disrupted form of the gene, and remain hybridized under the conditions use. Useful probes include, but are not limited to, nucleotide sequences which distinguish between a gene described herein and an altered form of that gene shown to be associated with reduced sperm count (azoospermia, oligospermia). Generally, the probe will be at least 7 nucleotides, while the upper limit is the length of the gene itself, e.g., up to about 40,000 nucleotides in length. Probes can be, for example, 10 to 14 nucleotides or longer (e.g., 20, 30, 50, 100, 250 nucleotides or any other useful length); the length of a specific probe will be determined by the assay in which it is used.

In one embodiment, the present invention is a method of diagnosing or aiding in the diagnosis of reduced sperm count associated with deletion or alteration of a gene described herein. Any man may be assessed with this method of diagnosis. In general, the man will have been at least preliminarily assessed, by another method, as having a reduced sperm count. By combining nucleic acid probes derived either from the isolated native sequence or cDNA sequence of the gene, or from appropriate primers, with the DNA from a sample to be assessed, under conditions suitable for hybridization of the probes with unaltered complementary nucleotide sequences in the sample but not with altered complementary nucleotide sequences, it can be determined whether the man possesses the intact gene. If the gene is unaltered, it may be concluded that the alteration of the gene is not responsible for the reduced

-24-

sperm count. This invention may also be used in a similar method wherein the hybridization conditions are such that the probes will hybridize only with altered DNA and not with unaltered sequences. The hybridized DNA can also be
5 isolated and sequenced to determine the precise nature of the alteration associated with the reduced sperm count. DNA assessed by the present method can be obtained from a variety of tissues and body fluids, such as blood or semen. In one embodiment, the above methods are carried out on DNA
10 obtained from a blood sample.

The invention also provides expression vectors containing a nucleotide (nucleic acid) sequence described herein, which is operably linked to at least one regulatory sequence. "Operably linked" is intended to mean that the
15 nucleotide sequence is linked to a regulatory sequence in a manner which allows expression of the nucleotide sequence. The term "regulatory sequence" included promoters, enhancers, and other expression control elements (see, e.g., Goeddel, Gene Expression Technology: Methods in
20 Enzymology 185, Academic Press, San Diego, CA (1990)). It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the protein or peptide desired to be expressed. For instance, the peptides of the
25 present invention can be produced by ligating the cloned gene, or a portion thereof, into a vector suitable for expression in either prokaryotic cells, eukaryotic cells or both (see, for example, Broach, et al., Experimental Manipulation of Gene Expression, ed. M. Inouye (Academic
30 Press, 1983) p. 83; Molecular Cloning: A Laboratory Manual, 2nd Ed., ed. Sambrook et al. (Cold Spring Harbor Laboratory Press, 1989) Chapters 16 and 17).

Prokaryotic and eukaryotic host cells transfected by the described vectors are also provided by this invention.
35 For instance, cells which can be transfected with the

-25-

vectors of the present invention include, but are not limited to, bacterial cells such as *E. coli*, insect cells (baculovirus), yeast and mammalian cells, such as Chinese hamster ovary cells (CHO).

5 Thus, a nucleotide sequence described herein can be used to produce a recombinant form of the protein via microbial or eukaryotic cellular processes. Production of a recombinant form of the protein can be carried out using known techniques, such as by ligating the oligonucleotide
10 sequence into a DNA or RNA construct, such as an expression vector, and transforming or transfecting the construct into host cells, either eukaryotic (yeast, avian, insect or mammalian) or prokaryotic (bacterial cells). Similar procedures, or modifications thereof, can be employed to
15 prepare recombinant proteins according to the present invention by microbial means or tissue-culture technology.

 The present invention also pertains to pharmaceutical compositions comprising the proteins and peptides described herein. For instance, the peptides or proteins of the
20 present invention can be formulated with a physiologically acceptable medium to prepare a pharmaceutical composition. The particular physiological medium may include, but is not limited to, water, buffered saline, polyols (e.g., glycerol, propylene glycol, liquid polyethylene glycol) and
25 dextrose solutions. The optimum concentration of the active ingredient(s) in the chosen medium can be determined empirically, according to procedures well known to medicinal chemists, and will depend on the ultimate pharmaceutical formulation desired. Methods of
30 introduction of exogenous polypeptides at the site of treatment include, but are not limited to, intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, oral and intranasal. Other suitable methods of introduction can also include rechargeable or biodegradable
35 devices and slow release polymeric devices. The

-26-

pharmaceutical compositions of this invention can also be administered as part of a combinatorial therapy with other agents.

This invention also has utility in methods of treating disorders of reduced sperm count associated with deletion or alteration of a gene described herein. These genes may be used in a method of gene therapy, whereby the gene or a gene portion encoding a functional protein is inserted into cells in which the functional protein is expressed and from which it is generally secreted to remedy the deficiency caused by the defect in the native gene.

The present invention is also related to antibodies which bind a protein or peptide encoded by all or a portion of a gene of the present invention, as well as antibodies which bind the protein or peptide encoded by all or a portion of a disrupted form of the gene. For instance, polyclonal and monoclonal antibodies which bind to the described polypeptide or protein are within the scope of the invention. A mammal, such as a mouse, hamster or rabbit, can be immunized with an immunogenic form of the protein or peptide (an antigenic fragment of the protein or peptide which is capable of eliciting an antibody response). Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques are well known in the art. The protein or peptide can be administered in the presence of an adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibody.

Following immunization, anti-peptide antisera can be obtained, and if desired, polyclonal antibodies can be isolated from the serum. Monoclonal antibodies can be isolated from the serum. Monoclonal antibodies can also be produced by standard techniques which are well known in the

-27-

art (Koehler and Milstein, Nature 256: 495-497 (19775); Kozbar et al., Immunology Today 4: 72 (1983); and Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96 (1985)). Such antibodies are useful
5 as diagnostics for the intact or disrupted gene and also as research tools for identifying either the intact or disrupted gene.

The present invention is illustrated by the following examples, which are not intended to be limiting in any way.

10 EXAMPLE 1 ISOLATION OF CDNA CLONES FROM HUMAN TESTIS
LIBRARY

"cDNA selection" (M. Lovett et al., *Proc. Natl. Acad. Sci. USA* 88:9628 (1991)) was carried out using bulk cDNA prepared from human adult testes (Clontech, Palo Alto, CA)
15 and, as selector, a cosmid library prepared from flow-sorted Y chromosomes (Lawrence Livermore National Laboratory: LL0YNC03). A total of 3600 random cosmids, providing nearly five-fold coverage of the 30-Mb euchromatic region, were used to generate 150 pools of
20 selector DNA. Using each of the 150 selector pools, we carried out four successive rounds of cDNA selection, followed by two rounds of subtraction with human COT-1 DNA (Gibco BRL, Gaithersburg, MD) to remove highly repetitive sequences. A plasmid library was prepared from each of the
25 150 resulting pools of selected cDNA fragments, and 24 clones from each library were sequenced from one end. Of the 3600 sequences generated, about 600 were of poor technical quality and about 500 were found to derive from cloning vector or *E. coli* host, leaving 2539 sequences for
30 further analysis. Of the 2539 sequence fragments, 536 corresponded to previously reported NRY genes (487 to *TSPY*, 15 to *YRRM*, 14 to *RPS4Y*, 9 to *SMCY*, 5 to *DAZ*, 3 to *SRY*, 3 to *ZFY*) and 41 corresponded to previously reported pseudoautosomal genes (15 to *XE7*, 11 to *CSF2RA*, 4 to *IL3RA*,

-28-

3 to ASMT, 3 to IL9R, 2 to ANT3, 2 to MIC2, 1 to SYBL1). Electronic analysis of the roughly 2000 remaining sequences revealed that about 200 contained known repetitive elements, and these were not pursued. By electronically
5 identifying redundancies and sequence overlaps, the remaining sequences were reduced to 1093 sequence contigs. Sequences representing these 1093 contigs were individually hybridized to dot-blotted yeast genomic DNAs of 60 YACs comprising most of the Y's euchromatic region (S. Foote et
10 al., *Science* 258:60 (1992)). 181 sequences that hybridized to the great majority of the YACs were judged likely to contain highly repeated elements and were not pursued, leaving 912 sequences for further analysis. The 912
15 sequences were individually hybridized to Southern blots of R1-digested human 46,XX female and 49,XYYYY male (L. Sirota et al., *Clin. Genet.* 19:87 (1981)) genomic DNAs. Blots were hybridized at 65°C in Church's buffer (0.5 M Na₂PO₄ at pH7.5, with 7% SDS), and washed at 65°C in 1X SSC and 0.1% SDS, with 832 hybridizations yielding interpretable
20 results. Many sequences appeared to contain highly repeated elements common to males and females, or failed to detect an unambiguously Y-specific restriction fragment, and these were not pursued. By contrast, 308 sequences hybridized to at least one prominent fragment present in
25 49,XYYYY but absent in 46,XX, suggesting that these sequences derived from the NRY. Each of these 308 sequences was individually used to screen, by hybridization, about 2 million plaques from a 1 phage library of human adult testis cDNA (Clontech, Palo Alto,
30 CA).

EXAMPLE 2 LOCALIZATION OF 12 NOVEL GENES ON THE Y CHROMOSOME

Genes were localized on a previously reported NRY deletion map by testing with PCR for their presence or

-29-

absence in individuals carrying partial Y chromosomes (D. Vollrath et al., *Science* 258:52 (1992)). Most genes were localized to a single deletion interval. Some genes could not be unambiguously placed because copies exist in

5 multiple locations in the NRY. In such cases, genes were localized by PCR testing of YACs encompassing the NRY's euchromatic region (S. Foote et al., *Science* 258:60 (1992)). X homologs of Y genes were mapped onto the X by

10 PCR testing a panel of human/rodent somatic hybrid cell lines (Research Genetics, Huntsville, AL). All PCR assays consists of 30 cycles of the following conditions: 1 min denaturing at 94°C, 45 sec annealing at 60°C, and 45 sec extension at 72°C. TB4X primers were designed from an unreported intron. TPRX primers were designed from

15 unreported cDNA sequence. All other primers were designed from cDNA sequences as submitted to Genbank. PCR primers were as follows:

GENE	LEFT PRIMER	RIGHT PRIMER
DBY	CATTCGGTTTTACCAGCCAG	CAGTGACTCGAGGTTCAATG
20	(SEQ ID NO.: 51)	(SEQ ID NO.: 52)
TPRY	GCATCATAATATGGATCTAGTAGG	GGAGATACTGAATAGCATAGC
	(SEQ ID NO.: 53)	(SEQ ID NO.: 54)
TB4Y	CAAAGACCTGCTGACAATGG	CTCCGCTAAGTCTTTCACC
	(SEQ ID NO.: 55)	(SEQ ID NO.: 56)
25 EIF1AY	CTCTGTAGCCAGCCTCTTC	GACTCCTTTCTGGCGGTTAC
	(SEQ ID NO.: 57)	(SEQ ID NO.: 58)
DDFRY	GAGCCCATCTTTGTCAGTTTAC	CTGCCAATTTTCCACATCAACC
	(SEQ ID NO.: 59)	(SEQ ID NO.: 60)
CDY	GGCTCAAATCCACTGACG	CAAGCGATATCTCACCACC
30	(SEQ ID NO.: 61)	(SEQ ID NO.: 62)
BPY1	CTCCCTGAGCAGCAACTAAG	GTCATCAACATGGGAAGCAC
	(SEQ ID NO.: 63)	(SEQ ID NO.: 64)
BPY2	CCAGGACCATGTGATATGG	CTAATTCCTCTTTACGCATGACC
	(SEQ ID NO.: 65)	(SEQ ID NO.: 66)

-30-

XKRY	CACTCATGGAGAAGGGTAGG (SEQ ID NO.: 67)	GTCACACTCAGCCTCTTTAC (SEQ ID NO.: 68)
PTPRY	GAGCACACCACACCAGAAAC (SEQ ID NO.: 69)	CTCAGACTGACCTCGGACTG (SEQ ID NO.: 70)
5 TTY1	CTCTGGGAATCAAATTCGAGG (SEQ ID NO.: 71)	GTCTTTCAGCCAATCCAAGG (SEQ ID NO.: 72)
TTY2	GACAACTCTGACAGCCAGG (SEQ ID NO.: 73)	GTCAGAACTCCCAAACAGG (SEQ ID NO.: 74)
DBX	CTACATGCAGATGACATGGTG (SEQ ID NO.: 75)	GGCCAAGGTGCATAGGTG (SEQ ID NO.: 76)
10 TPRX	CATGTTCCCTGTAGCACATC (SEQ ID NO.: 77)	CGTTTCCATTACTTCCATTTCTG (SEQ ID NO.: 78)
TB4X	CCCGCCCTTTCATCATCC (SEQ ID NO.: 79)	GCTCCCCAAAGTAGCCTTC (SEQ ID NO.: 80)
15 EIF1AX	CACGAGGCGCCATTTGCTG (SEQ ID NO.: 81)	CTGGAGGCCAGGCAACGTG (SEQ ID NO.: 82)
DFFRX	CCTCCACCTGAAGATGCC (SEQ ID NO.: 83)	CTGAGATCCAGGTGAATGG (SEQ ID NO.: 84)

EQUIVALENTS

20 Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

-31-

CLAIMS

We claim:

1. Isolated testis-specific DNA which occurs on the non-recombining region of the human Y chromosome or the complement thereof.
5
2. The isolated testis-specific DNA of Claim 1 which occurs in multiple copies on the non-recombining region of the human Y chromosome or the complement thereof.
3. The isolated testis-specific DNA of Claim 2 selected
10 from the group consisting of:
 - (a) a CDY gene or a characteristic portion thereof;
 - (b) a BPY 1 gene or a characteristic portion thereof;
 - (c) a BPY 2 gene or a characteristic portion thereof;
 - (d) an XKRY gene or a characteristic portion thereof;
 - 15 (e) a PTPRY gene or a characteristic portion thereof;
 - (f) TTY1 DNA; or a characteristic portion thereof;
 - (g) TTY 2 DNA; or a characteristic portion thereof;
 - (h) a complement of (a);
 - (i) a complement of (b);
 - 20 (j) a complement of (c);
 - (k) a complement of (d);
 - (l) a complement of (e);
 - (m) a complement of (f);
 - (n) a complement of (g);
 - 25 (o) DNA encoding the amino acid sequence of SEQ ID No.: 39;.
 - (p) DNA encoding the amino acid sequence of SEQ ID No.: 40;
 - (q) DNA encoding the amino acid sequence of SEQ ID No.: 42;
 - 30 (r) DNA encoding the amino acid sequence of SEQ ID No.: 44;

-32-

- (s) DNA encoding the amino acid sequence of SEQ ID No.: 46;
 - (t) DNA encoding the amino acid sequence of SEQ ID No.: 48; and
 - 5 (u) DNA which hybridizes to a DNA of any one of (a) through (t) under stringent conditions.
4. Isolated testis specific DNA selected from the group consisting of:
- (a) DNA of SEQ ID No.: 37;
 - 10 (b) DNA of SEQ ID No.: 38;
 - (c) DNA of SEQ ID No.: 41;
 - (d) DNA of SEQ ID No.: 43;
 - (e) DNA of SEQ ID No.: 45;
 - (f) DNA of SEQ ID No.: 47;
 - 15 (g) DNA of SEQ ID No.: 49;
 - (h) DNA of SEQ ID No.: 50;
 - (i) DNA encoding the amino acid sequence of SEQ ID No.39;
 - (j) DNA encoding the amino acid sequence of SEQ ID No.40;
 - 20 (k) DNA encoding the amino acid sequence of SEQ ID No.42;
 - (l) DNA encoding the amino acid sequence of SEQ ID No.44;
 - 25 (m) DNA encoding the amino acid sequence of SEQ ID No.46;
 - (n) DNA encoding the amino acid sequence of SEQ ID No.48;
 - (o) a complement of a DNA of any one of (a) through (n); and
 - 30 (p) DNA which hybridizes to a DNA of any one of (a) through (o) under stringent conditions.

-33-

5. Isolated X-homologous DNA which occurs on the non-recombining region of the human Y chromosome, is not testis-specific and has a homolog on the human X chromosome.
- 5 6. The isolated DNA of Claim 5 selected from the group consisting of:
- (a) a DBY gene or a characteristic portion thereof;
 - (b) a TPRY gene or a characteristic portion thereof;
 - 10 (c) a TB4Y gene or a characteristic portion thereof;
 - (d) an EIF1AY gene or a characteristic portion thereof;
 - (e) a DFFRY gene or a characteristic portion thereof;
 - 15 (f) a complement of (a);
 - (g) a complement of (b);
 - (h) a complement of (c);
 - (i) a complement of (d);
 - (j) a complement of (e);
 - 20 (k) a complement of (f);
 - (l) DNA encoding the amino acid sequence of SEQ ID No.: 18;
 - (m) DNA encoding the amino acid sequence of SEQ ID No.: 22;
 - 25 (n) DNA encoding the amino acid sequence of SEQ ID No.: 23
 - (o) DNA encoding the amino acid sequence of SEQ ID No.: 24;
 - (p) DNA encoding the amino acid sequence of SEQ ID No.: 28;
 - 30 (q) DNA encoding the amino acid sequence of SEQ ID No.: 32;
 - (r) DNA encoding the amino acid sequence of SEQ ID No.: 36; and;

-34-

- (s) DNA which hybridizes to a DNA of any one of (a) through (r) under stringent conditions.

7. Isolated X-homologous human DNA selected from the group consisting of:

- 5 (a) DNA of SEQ ID No.: 17 or a characteristic portion thereof;
- (b) DNA of SEQ ID No.: 19 or a characteristic portion thereof;
- 10 (c) DNA of SEQ ID No.: 20 or a characteristic portion thereof;
- (d) DNA of SEQ ID No.: 21 or a characteristic portion thereof;
- (e) DNA of SEQ ID No.: 26 or a characteristic portion thereof;
- 15 (f) DNA of SEQ ID No.: 30 or a characteristic portion thereof;
- (g) DNA of SEQ ID No.: 34 or a characteristic portion thereof;
- (h) DNA encoding the amino acid sequence of SEQ ID
- 20 No.: 18;
- (i) DNA encoding the amino acid sequence of SEQ ID No.: 22;
- (j) DNA encoding the amino acid sequence of SEQ ID No.: 23;
- 25 (k) DNA encoding the amino acid sequence of SEQ ID No.: 24;
- (l) DNA encoding the amino acid sequence of SEQ ID No.: 28;
- (m) DNA encoding the amino acid sequence of SEQ ID
- 30 No.: 32;
- (n) DNA encoding the amino acid sequence of SEQ ID No.: 36;
- (o) a complement of a DNA of any one of (a) through (n); and

-35-

(p) DNA which hybridizes to a DNA any one of (a) through (o) under stringent conditions.

8. A DNA probe comprising all or a characteristic portion of DNA of Claim 4.
- 5 9. A DNA probe comprising all or a characteristic portion of DNA of Claim 7.

1/17

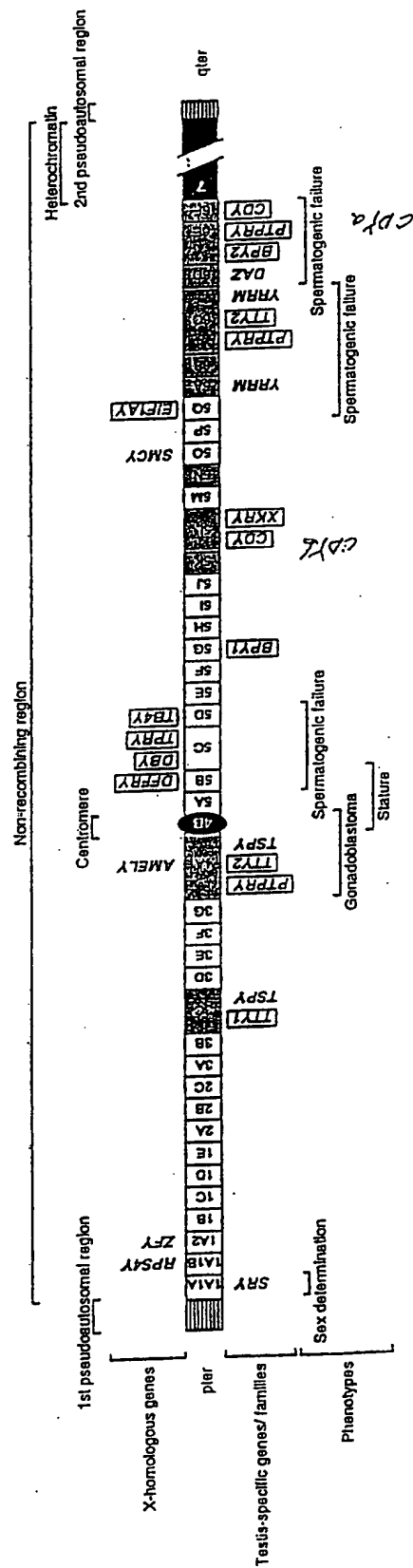


FIG. 1

2/17

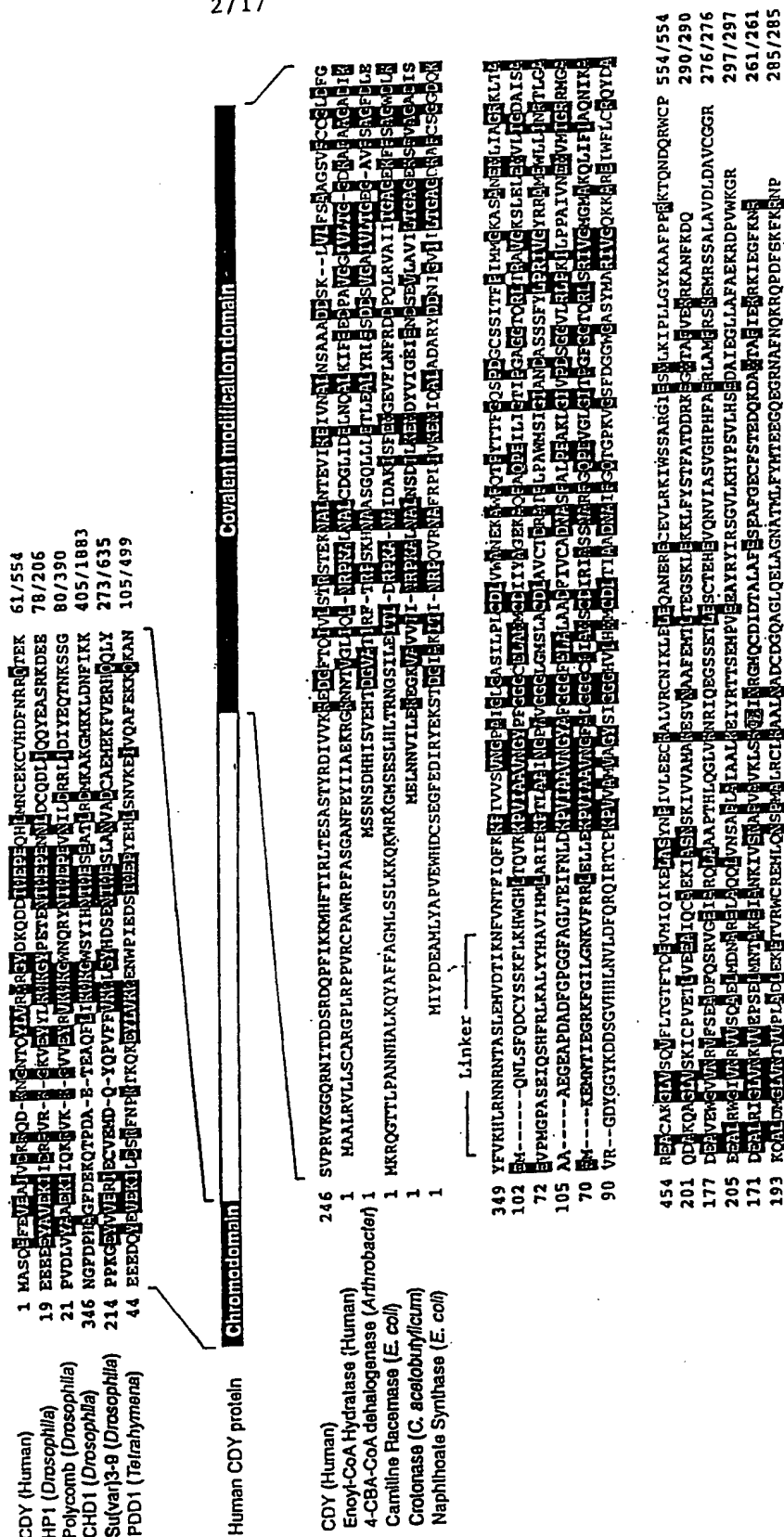


FIG. 2

long and ^{DBX & DBY} short transcripts

[illegible]

FIG. 3A

FIG. 3B

7/17

TB4X & TB4Y

[illegible]

FIG. 5

EIF1AX & EIF1AY

[illegible]

FIG. 6

FIG. 7B

10/17

1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010

FIG. 7C

FIG. 7D

12/17

CDYa & CDYb

CDYa -3281
CDYb -328 gtaacaggcaggaagaaagctttctgtactacaccagagggttggggctgtggattta
CDYa -270
CDYb -270 gctactctcacctgaggctactgagcaagtgtcatgcaccatgagacaaagcccaagctgtccaccaggcagtaagtatggagaggtt
CDYa -180
CDYb -180 caggcacatggcatagctgctatcttcgcacaattttcactacaccagtggtgacaaaatagaagaggttcatccatcacagaaacctggt
CDYa -90
CDYb -90 gaagagctggaggcagaagaagtgtctatgtggagacgcaactgaacaaaggtggcacagcaactgttccaatcccgtgtcttctcctc
1 M A S O E F E V E A I V D K R O D K N G N T O Y L V R W K G
CDYa 1
CDYb 1 ATGGCTTCCCAGGAGTTTGAAGCTATTGTTGACAAAAGACAGGATAAAAATGGGAATACACAGTATTGCTTGGTGGAAAGGT
31 Y D K O D D T W E P E O H L M N C E K C V H D F N R R O T E
CDYa 31
CDYb 31 TATGACAAACAGGATGACACTTGGGAACAGAGCAGCAGCTCATGAACCTGTGAAAAATGTGTACATGATTTTATAGACGACAGACTGAA
61 K O K K L T W T T T S R I F S N N A R R R R T S R S T K A N Y
CDYa 61
CDYb 61 AAACAGAAAACTGACATGGACTACAACCACTAGATTTTTCACAAATGCCAGAAGAAGACTTCCAGATCTACAAAAGCAACTAT
91 S K A N S P K T P V T D K H H R S K N R K L F A A S K N V R R
CDYa 91
CDYb 91 TCTAAGAACTCTCTAAACGCCAGCTGACTGATAACACCACAGGTCCAAAACCGCAAGTTATTGCTGCGCAGCAAGAAGCTTAGGAGA
121 K A A S I L S D T K N M E I I N S T I E T L A P D S P F D H
CDYa 121
CDYb 121 AAGGAGCTTCAATTCTCTCCGACACAAAGATATGGAGATAATAATCAACTATGAGACCTTGACCTGACAGCCCTTTGACAC
151 K T V S G F O K L E K L N P I A A D O O D T V V F K V T E
CDYa 151
CDYb 151 AAA---ACTGTGAGTGGCTTTCAGAACTTGAGAACTGAACCTATTGCGAGATCAGCAGGACAGCGTGGTCTTCAAGGTGACAGAA
180 G K L L R D P L S R P G A E O T G I O N K T O I H P L M S O
CDYa 180
CDYb 180 GGGAACTCTCCGGGACCTTTGTCACTGCTGGTGCAGAACAGACTGGAAATACAGAACAGACTGACACCCCTTAATGTGCGAG
210 M S G S V T A S M A T G S A T R K G I V V L I D P L A A N G
CDYa 210
CDYb 210 ATGCTAGCTCAGTTACTGCTTCTATGGCCACAGGTTGAGTACCCGAAAGGGTATAGTGGTATTATAGACCCATTAGCAGCCAAATGGG
240 T T D M H T S V P R V K G G O R N I T D D S R D O P F I K K
CDYa 240
CDYb 240 ACACAGACATGCATACCTCAGTTCCAAGAGTGAAGGTGGGCAAGAAATATTACTGATGACAGAGACAGCGCTTCTTATGAGAAAG
270 M H F T I R L T E S A S T Y R D I V V K K E D G F T O I V L
CDYa 270
CDYb 270 ATGCACTTACCATAAGGCTAACAGAAAGTGCACGACATACAGAGACATTGTAGTGAAGAAAGAGGATGGATTACCCAGATAGTGCTA
300 S T R S T E K N A L N T E V I K E I V N A L N S A A A D D S
CDYa 300
CDYb 300 TCACTAGATCGACAGAAAAATGCACTGAATACAGAAATTAAGAAATAGTTAATGCTCTGAATAGCGCTGCTGCAGATGACAGC
330 K L V L F S A A G S V F C C G L D F G Y F V K H L R N N R N
CDYa 330
CDYb 330 AAGCTCGTGTGTTTCACTGCGAGCTGGAAGTCTTTTCTGCTGGGCTTGATTTGGGTACTTTGTAAGCACTTAAGGAATAACAGAAAC
360 T A S L E M V D T I K N F V N T F I O F K K P I V V S V N G
CDYa 360
CDYb 360 ACAGCAAGCTTGAATGGTGGACACCATCAAGAACTTTGTAATATCTTTTATCAATTTAAAGCCTATTGTTGTATCAGTCAATGGC
390 P A I G L G A S I L P L C D L V W A N E K A W F O T P Y T T
CDYa 390
CDYb 390 CCTGCGATGGACTAGGTGCATCCATCCTGCTCTTGTGATCTCGTGGGCTAATGAAAGGCTTGGTTCCAAACCCCTTATACGACC
420 F G O S P D G C S S I T F P I H M G K A S A N E M L I A G R
CDYa 420
CDYb 420 TTTGACAGAGTCCAGATGGCTGTTCTTCTATTACATCCCCATAATGATGGGTAAAGCATCTGCCAATGAATGTTAATGCTGGGCGA
450 K L T A R E A C A K G L V S Q V F L T G T F T O E V M I O I
CDYa 450
CDYb 450 AAGCTGACAGCAAGGAGGATGCGCCAAAGGCTGCTCTCAGGTATTTTGTGAGTGAACCTTACCCCAAGAGGTATGATTCAAATT
480 K E L A S Y N P I V L E E C K A L V R C N I K L E L E O A N
CDYa 480
CDYb 480 AAGAGCTTGCTCATACAATCCAAATGTACTGGAAGAATGTAGGCCCTCGTTCGCTGTAATATTAAGTTGGAGTTGGAACAGGCCAAT
510 E R E C E V L R K I W S S A R G I E S M L K I P L L G Y K A
CDYa 510
CDYb 510 GAGAGAGAGTGTGAGGTGTGAGGAAGATCTGGAGCTCAGCCGAGGATAGAAATCCATGTTAAAAATACCTCTGTGGGATATAAGCA
540 A F P P R K K T O N D O R W C P 554
CDYa 540
CDYb 540 GCCTTCCCTCCAGAAAGACACAGAAATGATCAGAGATGGTGCCTTGACTTctatagtggcacaaacgcttcagagacacacattataag
1708 agacttatcttttagcataaatacttatggctcaaaatccactgacgacatcttctctaaactgaacacatgactagaattggtggtgag
CDYa 1798
CDYb 1798 atatcgcttgattttctttctctataaagtctagtcttaccaggttaacaaagaaactttatcgctctaaagttaagactgtta
1888 caccacaaaaaa 1903

FIG. 8

BPY1

```

-72      gagagggggtatatacaggggaggccaggcagcctggagttagtcgaccgttgcgagacgcttgagctgcggcag
1  ATGAGTCCAAAGCCGAGAGCCTCGGGACCTCGGGCCAAGGCCAAGGAGACAGGAAAGAGGAAGTCCTCCTCTCAGCCGAGGCCCAAGTGGC
1  M S P K P R A S G G P P A K A K E T G G K R S S S S P S F S G
91  CGGAAGAAGAACTACCAAGTGGCCGAGAAGGGAAGACAGTCTCTGGAGAGACCGCGGGAAGAAAGGGGTCTCGCAAAAGATGGCG
31  P K K K K T T T K V A E K G E A V R G G R R G K K G A A T K M A
181 GCGCTGACGGCAGCTTAGCGGAGAGCGGGCGACCGGCGACCGGCCAGCGACCGCCAGCCAGGAGCTCCCTCAGCAGCAGCTGCGG
61  A V T A P T A G E A G S G P A A P G P S D Q P S Q E E L P Q H E E L C P
271 CGGAGGAGGCAGTGAAGAGGGGACCCAGCAGCCCTGAGTCAGAGAGCGGAGCTGGAGGAACCACTAGTAAGAGGGGCGCCATCT
91  P E E T P V S E G T Q H D P L S Q E S E L E E P L S K G R P S
361 ACTCCCTATCTCCCTGAGcgaactaagtttagggcccagctgccagacctcagagatctcaccagcaggggtgcttccccatgttgatga
121 T P L S F 125
451 caataaaatgaatgtgtgtgcgaataaaaaa 480
          94

```

FIG. 9

BPY2

```

-332      aatatcttcaggaccaggacatgtgatatgggcccacacatctgtagtgatgttactctctctg
-270 cctagggtcatgcgtaaaagaggaattagggcataattgcttggccacagtcataatgatatgtactctctctgtgtgcagagccacaga
-180 agtgtgcttggtagacataactcttggagctgcataccaggatctatctgatactctgagccagcatabaagctgacactcttgacta
-90  tggccagccttcaataatactacactgtatataattgggtccaacccaggatgataattgttccatttcaactgagaccagataaagaagccta

1  ATGATGACGGTCTGTGCCAGAGCCAGGACATCGTCAGGACAGGATCATTACTCTCATCCCGCCAGGATTTTCACAGGTCGCTTACA
   M H M T L G T G T C C C A G A G C C A G G A C A T C G T C A G G A C A G G A T C A T T A C T C T C A T C C C G C C A G G A T T T C A C A G G T C G C T T A C A
91  GAGGGCATGATGACATAATTGCTTGACAAAGAACCTAAAGTAGTGTTAAATATTTCGTACATAGGTGTGTAAANAATGGGAATGTGAGAAATACC
31  G G G G I M T T Y C L T K N A C S D V I L H R G R L L K A N G N V R N T
181 TTGCTCATGCTCAAAAGTGGGCTTGCTGCATATATTATGTGAAACCTGATCCCGGTTGAGTGTACTTCTTGACTAGGCCAGCATCAAAATG
61  L L Q S K K V G L L T Y Y V K L Y C P G E V T C T L L T T L T R . P S I Q M H
271 AGATTATGCTGTATCATCGGCTCAGTGTGCGAAGCCGACGATCAGACAGAAATattgtgccatatgtggaacaagcagctaaagcaatagataa
91  R L C C C I I T G G S V S K P R R Q K * 106

361 catccatactgtggctctgctctcctcctcaaaagggaaatttcatatctgtctactgtgggacatccaccagatgatgtctctgccctcaaaaagaatttgt
451 gacataacgctgactgacaaaacatgggtatgtgaacctctctcttattctctggagttctgccaaaacagagggattatcacatatctgtcgggag
541 tccagcaccacaggtaaaatttcttcataataccacagctctcagataccatgacatgatcaactacatctcagtgagcccaaaagaggagagat
631 attttgattctcattgccattcttattgtggccacaagaagtgaattgtcttcctcatagtggatataaagttccacagagattatgacatctccca
721 cgcgtatgataaaaattgtgtagtatacaactgagctggtcataacagggaacagcaaaacaaatgctatctgtgattattgttgattcacaccagc
811 gcagcgcatactattcttcacaaagacaagaacctgtcaataaattctataaaccctccacaaaaaanaa 880

```

FIG. 10

XKRY

```

-663          attaaaaactctctgataaaattacctgaagtaca
-630  cacaacacaaaaacatgcccacacaaatcacttaattttctaaaacttttaattttctctctctagctactcttgattccatccacacagc
-540  aaaaactctggcagctccacttcagaaatttacttgtaactccacagcttatctccgattttctctgttaccaggaggtctaaaaacacagattta
-450  tatgtcatctccactctctattttacacgcgttaatttctctactttacacttaactttatataaaaaaagaaactaccttttcagaatctcaatt
-360  cacgcgaattttattgtgtcttaatttgagacttttctcttaggtgcgtgcacacactgtgaacgtcagatcaaaattctctctccaatttcca
-270  ttggttccagcttattttattttaagggaatgtgatatacatatttaacttaatttggtagtgggtgtattcacttattcttattttatagc
-180  tagtctcgatataatttactaaaattccctgataagaataaacaattttttttctttctttttttttttttgtatgtaaatatttttcgga
-90  agggaggtgggttggggagaataatatctttaacttggcaagtttaaaagagaaggtggccattactaatgaaaattattctctagcattttc

      1  ATGTTTATCTTTTAATAGCATTCCTGATGACATATTCCCTCTTATCAGTTGTGTAGTGCCATTCACTGCAATATACTGGCCATCCGCACT
      1  M F I I F . N S I A A D D I F P L I S C V G A I H C N G I L A I R T
91  GGCACACGACTTGTCTGCCATTAACTACAGGTGATAAAATTGATCTATCTCATGATAGGCATTGGTGTGNTGATTAATCACTCTAGTG
31  G N D F A A I K L Q V I K L I Y L M I W H S L V I I S P V V
181  A T C T G G C A C T C T T C C C T G C A S L T K Q A A C A G G G A C T T A C A C T T C T A T T A T A T T T G T A T T G T G C A C C A T G G C T G
61  A L C A F F S C L K Q A A C A G G G A C T T A C A C T T C T A T T A T A T T T G T A T T G T G C A C C A T G G C T G
271  G A G T T S K G A A S G T G G A A C T A C T T C C T A G C A C A A A A A T A T T C C A G C A T G G T G G T A G T A T G G A T G C T T A C T T A A T C A T G C T
91  E F F T S K G A A S G T H L F S N T K I I P A W H V S M D A Y L N H A
361  A G T A T C G T G C C A C T A A T T C C T G C T T G C A G C A G T G A A A C T G C A G L C A T T C A A T G A G G A A T T G A T A G A G A C A C A G G A T G G G A C A T A
121  S I C C H Q F F S C L S A V K L Q G L S N E E L I R D T R W D I
451  C A A T C T A C A C A G A T T C A G T T T T A G a a a a t g t g a t a a t a a t t g a t a t c a a g t t c c t t g g a g g g a a c g t t t t a c c g a a g t g t t
151  Q S Y T T D F S F . 159
541  g t g a c t c a a t a a t t g c c g t g t a g t t c a t c a a a a c c t a c a t a t t a g c c t t t g g c t t t a a g c t c c g c t t c t g t c a g t a t t t g c a a c c a a g g t
631  g g t c g g g g a a a g t a t t g c c a g g a g a t a c t g a a a a c t c c a a g a a g c t g a t a t t t g t a a g c a t c t g g a g a a a t c a g t t a a a a g a
721  a t a a a g a a a g c a g c t g a g a a t t a c t a c t a c a t c a t g g a a g g g t a g g a t a t t t c a a a g t g a g t a g t a c a a t c c a t a t a t a c t t
811  c a g a a a a a a a g a t a a a a g g c t g a g t g t g a c t t t a a a g a t a c t a c t g a a a a a t a t a a c a a c a a a a c c t t g g a a g t a g t t t c t a a t
901  a a a a t g a t t t t c t a a a a a a a a a a 925

```

FIG. 11

14/17

PTPRY

-182
-180
-90

gaagaggagcacaccacacagaaacagacatcttcagtggtttcactgtctcaaccttatctgcacagtcgaggtcagtcgtgagagag
cttctgagagacccaggatgaagggatgcagtgaggtcaagagcccaacctcttctcactgacacccacctcttaaggactcagaagagac

1 ATGAATAAAATGGGCTCAACAATCCCAAGAAGAACCACTCAAGGACATGGGAGCCACTGGGCTTGGCTTCTACTTCCCTGGAAACAA
1 M N K M H G L N N P K K N H S R T M G A T G L G F L L P W K Q
91 GACAATTGAATGGCACTGACTGCCAGGATGCAATATTTTACTTCTCTGAGACTACGGGGAGCATGTGTTCTGAACCTTCCCTGAAC
31 D N L N G T D C Q G C N I L Y F S E T T G S M C S E L S L N
181 AGAGGCTTGGAGCCAGAAGGAAGGATCTTAAAGACTACTTCTCTGGAGATATGGGAAGGTTGGCTGTATCTCACTTCCACTTCGT
61 R G L E A R R K K D L K D S F L W R Y G K V G C I S L P L R
271 GAGATGACCGCTGGATTAACCCACCCAAATTTTCAGAGTTCCTCAAGCTTACCACAGAGGTTGACCGAGCTGATCTGACTGAGCGCTG
91 E M T A W I N P F Q I S E I F Q G Y H Q R V H G A D A L S L
361 CAACCAACTCTCTGAGAAGCAGGTTATCTTACAGTGCCTCGGACAGAGCTTCTTCTCAGGACACTCGAGAGAGCGTGGTTTCAGGG
121 Q T N S L R S S Q C L G Q S F L L R T L E R A V V S G
451 CACTTGGGGACATCTGAGCCACGTTTCATGAAGAAGACTAAGCTTCTCTCAGGACCCGCCCAAGAGTGGCCGCGCTTGGGGACA
151 H L G T S V A T F M K K T K P T S S Q D P P K S G R G F G T
541 CCTGCGGTCCGCTCCACATGAGGATAAAACCTCTCTCTCTTGGACATGTCCAGGAGTGGCGGTGTCTACAAGTCACTGGTGCTACG
181 P A V G S T M R I K P P S L L D M S R S G R C Y K S P G A T
631 ACCAGGTTGAGAAATAAGACGTCTCTCAGGACCTCCAGGAGAGTACATGGCATTGAGACATCTGGCGCCCAAGTGAGGAAAAGACAC
211 T R V R I K T S P Q D P P R V H G I E T S G G Q V R K R H
721 CCTGCTGCAGCACCAGAACTGAGGAGGGGCACTGCGCTGGGCTTACTTCCAGCCCTGGCTCCAATTCTGACCTTACAAAAGTGTG
241 P V C S T Q N 247

811 ccttgagtcgagtcagtcacacgcattgtcacagctaccaaagtgtggtttgcagatgatctgggcttgtttctggcagagattctggta
901 cagagaaggagagggcggttgagtcggaaccagatgggctgagggccaggggagacatcacaaacctccaacaacactttttctatgcttta
991 ataaacatctttcttagagaactaaagttagtgaacaatatagaacaactttttaagtaggcat aaaaaaaa 1066

FIG. 12

TTY1

tgtctgtcagagctgtcagcctgcttaagcagagtaaaatggtacaggcagtcagcctggttagcgagaaaaaaggtgcctgtgaatc
ccactgtgggacccaatagtcgggacccctcagggccctctcatggcactccatggccatgtcatctgagagagggcggtttgaaagaa
tgagctgtctgctggaaactgtcatctgactccagctctcaaaagaggtcatgtgcaagaaatcggtgaaagtgtgagagcccatccacc
ctccacaagatgtcatccccacctgtctgaccttaactgtctcaaaactatctgtccaaggatgaaaaacccagagcaaaagggaggtaa
ccctcatgagtcgaagcagctgttcacctgtgaatataacctgaggatcatgagactatctgtggtattccagagagaagacagacgagaa
gacacccgtgacattctccacggaggtctcctcttccacaaagatgcagatgcttcttgcagagactatcctgtgaaatcccaagagaaa
gacaggtgtggttccaaatgcgggtgcacactccagggaaatctcctctctcaccagctccagggcctctgccaatgtcatgagactattt
gtggattccacagagagaagataggtgaaaggtacagcatggcatccacacctccacagaggggtatccccacctctctgaccttattacc
ttattgtcttcaaaagtctctatccacagactgaaatcccaagacaatgggaagtccccctgatgagtggaagcacaactcctctgtgg
aatcaaatccagggtaaatataagcccggttagagatgaaatgatgtagtctctcctctggattgggtgaaagacaatataaacactgtgca
tatttctgttcaaaaaaa

FIG. 13

TTY2

aggcttgccatcaccacagatggcctctgagacactgttgaaccacatctgcacctgtgagagggccagtttgaggtatgagaacactgt
tccaatctgggacttgcctttgtcttgggttctgctcttccacagatggcaccctacccacccaggtatgaatgagtcgagagaggtcaagt
ccaggcatcttctgtgacaccccttctgggtatttcaggatataagttccatccatccaaagactgtccacatctcaccagaaatatttt
caatccctcatggggcatgattcttccaaaaaccccttccaggaatggagtcagagagtagtttccagagacaacctcaccagctcttga
acggctctgctccatgctgactgtgacatggagatggcatataagggccctaaagttaggacttttagggtaactgcaatgcgttatcac
aggcagcctttatctgtatcaaaagccagctctgctgtacatttctctcgtctaggcagggctgacagccctgacacctgggtgctcc
agtttgagtcactatgtggtatgtgctacttctggggcaatggacctgagctgtgagctgttagctgttcacaaatgaatgccagctt
gctagtcaaatctcccttgggttggtagagaaggggacctctgtggaggtacaaatgggtgtgactgtcactgtctctctgtgggt
ccatgggacagttccatgatctcagagaggggttagatgtgagccagcctgaagaaatgtcaagcagagcccccaggaatgaagcacaanaat
cactacagatcccaaaaggtctgcagaaattgtcagggcctgctcagacattgtaggggttagtcttattgaatgtgtccactgtaatt
tccaaactcagccttctctgttctccagcagtttctctctccaggggtgggcttctgcagaatgacacagcctcagaagctactgggct
gtgtgttactgtgggaggtgttgcaggtgttggatgtcagcatgtgtgtgtggcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt
aaatgaattctgtggaatcaggaatcagcaatgactagttaagctgtctgtgacagccgggttccccatctgtctgcccctgccaanaaa
caggtactcttctacaaagagagagagacacaccccaagaaacagacatctccaggtgtgcatataaagcagccaacccacagaca
ctagcactctgtgtctgcatagccctttaatttacctaagaattcagttccacagccaagtaggtgcttcatgtctgaggggtgcaatctc
catcatcttgagatttcatgctgtgtacagagaggtgtgacagcaataaggtcagataggggtgagtatcaacctggtgaaggggtgagtg
ggctccgtacactcaccagcaaaaggggtgaanaatagatgacacagaaatgtgttccaaactccatccccacatctccataattgcaanaat
cagtcacaacacatggcctgtgtttaggtgggagtcctccaaactcaggaagaatttggagtcgaatctgtggccaactctggaanaactc
ctgggtttgaggggttttaactcctgtagtcgaatgggaagtggaatagattgtgctgggtgggttgtggcctccacatttctgtctctct
tactgacttccatgtcctcattgtgtgtagggtcttctgtgtctggctcaacatcttccacactaaactcttccctgttccacagaagacc
atcacataatgcattgtagagggcctgcaagggcaggaatgagggagacagtcaggtcaagagccagccatcttccactgagacccca
cttctgggttctcagggctgtgtgacaggtctgacagcccatcaggaagcctgcatactcttagacacaaggaactgagcttgggtccca
ctagcatcacatagaagggccacatttgcctagggaataggtctgtgacttgtgggataagaaactccgtggagccaatccaaagagaga
cactatctcttccatctcaccgggcttataatctaccctgaatttgaatccacagagaggtgtgtgtcttcaatcatcaggggggaactc
tccatgtctgggatttccagctctgggacagagacttgaacagcaataaggttctctgtgtctggctcaacgtcttctaaactcaacat
ccccagttcatggaaaatgatctcctcagggattctattgtaaaagctgcatgattctctgggagaaatgttttattgtcctggagtcaccca
aaaatcaacacatccagctcatgttgcagagctccttgaatcaacttgaatccagttccagctgagcagctgtctcagctgtgtgagggg
gcaatccctccatcaactcctgggtatttccattctgggacacagagatttgagcagcaataaggtgggtccacattgcccctcaacagcatag
tggaatgatgtgtcagacttgcatttccgcagacacactctctggaacatttttcaacatcatctacatgagtgagagaccggtctgaca
tgtaagaatactgcttgaacttggacctgccttctgtctgtgtctcctcctcttctcagatccctgccaggccagagatgagagggc
aatgaagtcagggtccgagcccatctcattgaaggtgactctggggctctgggttaattccatccacataaaatccccctcaacacactca
cagactatactccaaatctccatggaaactgtattcttgcacacagcctcttccaggaatggagtcagagagcagctcttccagagaccc
tcagtttggaaaagccctcctcagtggttccacagccatggagtcactctggaaggggtccacaggtcaaatcttccaggtactgac
ttgggttaccacagacagacttttccatgtagcatgcatctctctctatcttctcctgtctaggcaggtgcacacactgcagac
ccagggggccgaatctcctgggcaatgtgcatgctctactctcagtcgaaagggcgtgttgggagctctgactagtgtcacaaataaa
ggcccatgtgcttagtgaaaaaa

FIG. 14

Human CDYL (CDY Like)

ggagagaggacctatcttctacctaaggacattcccgggaaggcaatggggtttcaacaatat
cctgaagagactcatctcggggaactaagcaggtggtaatcagagaacacagagcccccgg
aagaattttatggcatttcagggaagccacaggccagcctggggaaaaagcaggaagaaaa
actggcaatacagagggcccaacccaaaagtattctcctgaagagaaacaacgtgtcagcacc
agatgggaccttcagaccccagcatctccgcgagcagtgagcaaagcggggcacagcagcct
ccccgtttacaggttgaaaggattggttgacaaaaggaaaaataaaaaagggaagacagagt
atcttggttcgggtggaaaggctatgacagcgaggacgacacttgggagccgggaacagcacct
cgtgaactgtgaggaatacatccacgacttcaacagacgccacacggagaagcagaaggag
agcacattgaccagaacaacaggacctctcccaacaatgctaggaaacaaaatctccagat
ccaccaacagcaacttttctaagacctctcctaaggcactcgtgattgggaaagaccacga
atccaaaaacagccagctgtttgctgcccagcagaagttcaggaagaacacagctccatct
ctctccagccggaagaacatggacctaagcgaaggttatcaagatcctcgtgcctaaaa
gccccgttaagagcaggacccgagtgaggcgttttcagagcgagagccctgagaaactgga
ccccgtcgagcagggtcaggaggacacagtggtcaccggaagtggcagcggaaaagccggtc
ggagcttttattggggccccgggtgcccagagggccaggatggggagcaggcccaggatacacc
cactagtgcctcaggtgccccggccctgtgactgcagccatggccacaggccttagctgttaa
cgggaaagggtacatctccgttcatggatgcattaacagccaatgggacaaccaacatacag
acatctgttacaggagtgactgccagcaaaaaggaaaatttattgacgacagaagagaccagc
cttttgacaagcgattgcgtttcagcgtgaggcaaacagaaaagtgcctacagatacagaga
tattgtggtcaggaagcaggatgggttcacccacatcttgttatccacaaagtcctcagag
aataactcactaaatccagaggtaatgagagaagtcagagtgctctgagcagggccgctg
ccgatgacagcaagctggtactgctcagcgcgcttggcagcgtcttctgtgtggtacttga
ctttatattttatacagcgtctgacagatgacaggaaaaagagaaaagcactaaaatggca
gaagctatcagaaacttcgtgaatactttcattcaatttaagaagccattattgtagcag
tcaatggcccagccattgggttcaaacaccctataaccaccttcggacagagtccagatggctgt
tctaccgttatgtttcccaagataatgggaggagcatctgcaaacgagatgctgctcagtg
gacggaagctgacagcgcaggaggcgtgtggcaaggccctgggtctcccaggtgttttggcc
cgggacgttcactcaggaagtgtggttcgcatttaaggagcttgccctcgtgcaatccagtt
gtgcttgaggaatccaaagccctcgtgcgctgcaacatgaagatggagctggagcaggcca
acgagaggggagtgtgaggtgctgaagaaaatctggggctcggcccaggggatggactccat
gttaaagtacttgacagaggaagatcgatgagttctgagtgctcgggctgcccactgggtgaca
ccgggatcgggctgagcaggagaacatcacgggtccagttcccctgatccattctcagag
cctgaaacaagctcacccgtagcttacgcttggaagcaggactgggaacatccacgctatt
tattatcgaggagttttaaagtactgtaactttaaaaataaataactacaaagcttcttgt
cvaacgctcattattttatacttatatacagcaggtgtaaaagtataaagggtgagcacta
gactgctcttagaagctctaatttttgggtttcttttggctagtactgtataaaaaacagaat
tgtgttttattgggttttggatgacagaaaagtctggaataatgtttgttttctcatttct
tcttcttagaacacagaatctaagggggtgttagccagcctcgccctccctgcccacgtag
agacacagagtgtgtgagggcgttggctttttctccaagaaggtacagatacctcagattc
gggaaactcaaaatcaaaagacttagcttctaggataaaatacttctgatgaaaaatccgct
gaggagcataaccccaaaccagacatatgcttaggattcatgctgagatatcaattgggttc
cccttcttttttaaaatacgtccagttcttaccaggttaacatgaagaaccactgtctcta
gaagaaagcttgttttgacgtatttagtgaatcactgaatagcttaagtatgactatctaag
ttataagttagtcttttagtgggttttaaatagtttttctgaccttctgaaaaataactac
ataagtgtcttcttgttgggtgagaaataactactttatagacagtttttgggtttctgtt
tgcagatatgattgatgtatttcacaaaaataaaatatttttatgtttataaagtgttaatt
tttaggttacttagaataatattttatttaataagttaaaattcttttggcacactattaa
atgcaaaaactcctttc

FIG. 15

Mouse *Cdyl* (CDY like)

ctttgaggtgggttagcatcccacttgttccttgaggacatctgttcctacctaagagcac
tcacctgagatgctcaaaggtccagaagaaacacttctcgggtgacaaagcaggtgggtgac
cagagaacagaggccccccaaaaattttatggcattcaaggcaaagcacagccaacccgga
gggaaagcaagagtccagcctggaaatacatagcccaacccgaaggttatctctgaaggaa
aacaatgggcataggcaatagccagcctaattcacaggaagcccagctctgcacacttcca
gagaaagctgaacaacactactgatgataaacacctgccagcaaaaataatgtgggtcctgcaa
cagtctcagaacccgatcaagcgtcccctgcaattcaagacgaggagactcaggtggaaag
tatcgttgacaaaaggaaaaacaagaaagggaagacagaatatctgggtgcggtggaaaggc
tatgacagtgaggatgacacgtgggagcctgagcagcacctgggtgaactgtgaggaataca
tccatgacttcaaccggcgccacaacgagaggcaaaaggaaggtagcctggctcgtgccag
cagagcctccccagcaacgcccgggaagcagatttccaggtccaccacagcactctctcc
aagaccaactccaaagcacttgtggtaggcaaaagatcatgagtccaaaagcagccagctgt
tggctgccagccagaagttcaggaaaaacccagccccatctcttgcaaacgcaagaacat
ggacctcgccaagtcagggatcaaaaattctcgtgcctaagagccccgttaagggcaggacc
tcgggtgatggctttcagggggagagccccgagaagctggaccctgtggatcaggggtgccg
aggacactgtagccccagaggtgactgcagagaagcccactggggctttgctgggcccctgg
tgccggagcagagccaggatggggagcaggccccgaatacatccactagtgcctcaggtttct
ggccccgtgactgctgccatggccacaggcttagctgttaatggaaaaggtacatctccat
tcatggatgctgtagcagccaacgggaacagtcaccatacacagacatccgtaacaggagtgc
agccgggaaaaaggaaatttattgacgacagagagaccaaccttttgacaagcgggttgctg
ttcaggtgtgaggcagacagagagtgccctacagatacacagagatatgtctcaggaagcaag
atggcttcccccacatctgttatccacaaaatcgtcagagaataactcactaaaccaga
ggatgaaagaagtrcagagcgccctgagcacagctgcagccgacgacagcaagctgggt
ctgctcagcgccgtgggcagcgtcttctgctgtgggtctggactttattttattttatccgc
gcctcacagatgaccgaaagagagaaagcactaaaatggcagacgctatcagaaactcgt
gaatactttcattcagtttaagaagcctattattgttagctgttaatggcccagccattgga
ctaggagcatccatattgcctctttgtgatgtgggtttgggctaacgaaaaggcttggtttc
aaacaccctataccaccttcggacagagtcagatgggtgctctaccgttatgtttcccaa
gattatggggaggagcatctgcgaatgaaatgctgttcagtgggcggaagttgacgggcacag
gaggcctgtggcaagggctgggtctcccaggtgttttggccaggaaccttcacacaggaag
tcatgggttcgaatcaaggagctggcttcatgtaacccagttgtcctggaggaatccaaagc
cctgggtgcgctgcaatatgaagatggagctagagcaggccaatgagagagaatgtgaagt
ctgaagaagatctggggctccgcccagggcatggactccatgttaaagtacttacagagga
aaatcgatgagttctgatgggcaggctgagcaggacatcggtgggtcccacttgctacgtc
gtcctgcagtggtcgtgcttggaggcgaaactggaaacatccgagctattttattgcccgcg
gagtttttaagtactgtaactttaaaaataaatacaaaagcttctttgtcctaagcgtctttat
ttatactcatgtatacacaagtataaaaaatgtaattgagcactaggctgctccttggaagc
tctaattttcttgtaagctagttgtggatttttgttttgtttttgttttttaaagggaatta
tgttttcattttgggtgacagaagagtttgaaataatgtttgtttttactctttttttttt
ccttaaatctagatcacagaccctcaaaattactagccagccttctccccctccctctact
gaaacatgtagaaataacttaaacatgttcctgcctctaggggggagggggaggtgtgagta
cctcaatgctgaaaacagttctgatcaaaacttaagaccaacctggtaaaaaaagcatcact
gatggaaaaatcccacccacgggggctgggtttctgctgaaatgcccgccgctctaccttt
cttactgtcccattcttaccagccaccgtgaagagcccagtgctcggaggaaagcaggtg
gtccagtgctgtgagtcactccgtagctcgagtggtacttgctaaagttagaattagcat
tagtgggtttaaatagtttttctgacccttttgaaaaataactacataagtaactcctgtg
ggctgggtgagaaataactactttgcatagtttttgtttgtctatctgcagatatgattgctg
tattacacccaaaagtattttttatgtttataaagtgtaatttttaggttcacttagaatat
attttatttaatttaaaattctcttggcacactattaaataacgtaaaactcctttc

FIG. 16

Human VCP (*Variably Charged Protein*) familyVCP2r (*VCP with 2 repeats*)

gttgcgagacgttgagctgcggaagatgagtcctaaagccgagagcctcgggacctccggcc
aaggccacggaggcaggaagaggaagtcctcctctcagccgagccccagtgacccgaaga
agaagactaccaaggtggccgagaagggaaaagcagttcgtagagggagacgcgggaagaa
aggggctgcgacaaagatggcggccgtgacggcacctgaggcggagagcgggcccagcggca
cccgccccagcagaccagcccagccaggagctccctcagcacgagctgccgcccggaggagc
cagtgagcagggggacccagcacgaccccccgagtcaggaggccgagctggaggaaccact
gagtcaggagagcaggtggaagaaccactgactgtgtggatggccagctttccctgtc
tccgagagcagcgactaagttcaggcccagccgagacctcagagatctcaccagcgggg
tgcttgccattctgaagataataaaatgaatgtgttgcaaattgaaaaaaaaa

FIG. 17A

VCP8r (*VCP with 8 repeats*)

cggaagatgagtcctaaagccgagagcctcgggacctccggccaaggccacggaggcaggaa
agaggaagtcctcctctcagccgagccccagtgacccgaagaagaagactaccaaggtggc
caagaagggaaaagcagttcgtagagggagacgcgggaagaaaggggctgcgacaaagatg
gcggccgtgacggcacctgaggcggagagcgggcccagcggcacccggccccagcagaccagc
ccagccaggagctccctcagcacgagctgccgcccggaggagccagtgagcagggggaccca
gcacgacccccctgagtcaggaggccgagctggaggaaccactgagtcaggagagcagaggtg
gaagaaccactgagtcaggagagccaggtggaggaaccactgagtcaggagagcagaggtgg
aggaaccgctgagtcaggagagccaggtggaggaaccactgagtcaggagagcagaggtgga
ggaaccactgagtcaggagagccaggtggaggaaccactgagtcaggagagcagagatggaa
gaactaccgagtgtagacggccagctactccctatctccgagagcagcgactaagttc
agggccagccgagacctcagagatctcaccagcggggtgcttgccattctgaagataat
aaaatgaatgtgttgcaaattgaaaaaaaaa

FIG. 17B

VCP10r (*VCP with 10 repeats*)

cgttgcgagacgttgagctgcggaagatgagtcctaaagccgagagcctcgggacctccggc
caaggccacggaggcaggaagaggaagtcctcctctcagccgagccccagtgacccgaag
aagaagactaccaaggtggccaagaagggaaaagcagttcgtagagggagacgcgggaaga
aaggggctgcgacaaagatggcggccgtgacggcacctgaggcggagagcgggcccagcggc
accgggccccagcagaccagcccagccaggagctccctcagcacgagctgccgcccggaggag
ccagtgagcagggggacccagcacgacccccctgagtcaggaggccgagctggaggaaccac
tgagtcaggagagcagaggtggaagaaccactgagtcaggagagccaggtggaggaaccact
gagtcaggagagcagaggtggaagaaccactgagtcaggagagccaggtggaggaaccactg
agtcaggagagcagaggtggaggaaccactgagtcaggagagccaggtggaggaaccactga
gtcaggagagcagagatggaagaaccactgagtcaggagagccaggtggaggaaccaccag
tcaggagagcagagatggaagaactaccgagtgtagacggccaagtactccctatctcc
gagagcagcgactaagttcaggcccagccgagacctcagagatctcaccagcggggtgc
ttgccattctgaagataataaaatgaatgtgttgcaaattgaaaaaaaaa

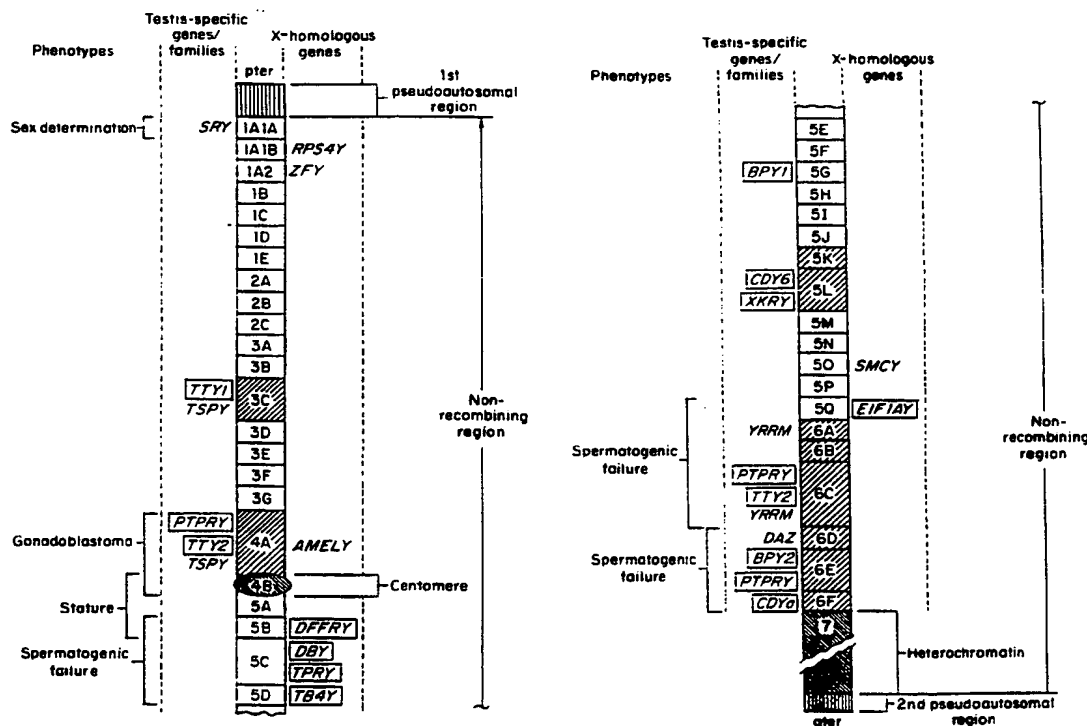
FIG. 17C



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/12, 15/11		A3	(11) International Publication Number: WO 98/46747
			(43) International Publication Date: 22 October 1998 (22.10.98)
(21) International Application Number: PCT/US98/07115		(81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(22) International Filing Date: 10 April 1998 (10.04.98)			
(30) Priority Data: 60/041,877 11 April 1997 (11.04.97) US		Published <i>With international search report.</i>	
(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 60/041,877 (CIP) Filed on 11 April 1997 (11.04.97)		(88) Date of publication of the international search report: 4 March 1999 (04.03.99)	
(71) Applicant (for all designated States except US): WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US).			
(72) Inventors; and (75) Inventors/Applicants (for US only): LAHN, Bruce, T. [CN/US]; 863 Massachusetts Avenue #26, Cambridge, MA 02139 (US). PAGE, David, C. [US/US]; 3 Ivy Circle, Winchester, MA 01890 (US).			
(74) Agents: GRANAHAH, Patricia et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02173 (US).			

(54) Title: GENES IN THE NON-RECOMBINING REGION OF THE Y CHROMOSOME



(57) Abstract

Genes of the non-recombining region of the human Y chromosome, which fall into two classes: X-homologous DNA which is expressed in many organs and has functional X homologs and testis-specific DNA.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

INTERNATIONAL SEARCH REPORT

International Application No.

P 98/07115

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C12N15/11

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 92 00375 A (IMP CANCER RES TECH) 9 January 1992 see the whole document ---	1
X	ZHANG J. ET AL.: "Molecular isolation and characterization of an expressed gene from the human Y chromosome" HUMAN MOLECULAR GENETICS, vol. 1, no. 9, December 1992, pages 717-726, XP002080218 see the whole document --- -/--	1,2

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

12 October 1998

Date of mailing of the international search report

17.12.98

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Kania, T

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/07115

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MA K. ET AL.: "A Y chromosome gene family with RNA-binding protein homology: candidates for the azoospermia factor AZF controlling human spermatogenesis" CELL, vol. 75, no. 7, 31 December 1993, pages 1287-1295, XP002017338 cited in the application see the whole document ---	1,2
X	WO 95 11300 A (MEDICAL RES COUNCIL ;CHANDLEY ANN CHESTER (GB); KUN MA (GB); SHARK) 27 April 1995 see the whole document ---	1,2
A	WO 97 10267 A (PROMEGA CORP ;KENT MARIJO G (US); AGULNIK ALEXANDER I (US)) 20 March 1997 see the whole document ---	1-4,8
A	PAGE D. ET AL.: "The sex-determining region of the human Y chromosome encodes a finger protein" CELL, vol. 51, no. 6, 24 December 1987, pages 1091-1104, XP002080219 cited in the application see the whole document ---	1-4,8
A	WO 96 41007 A (PROMEGA CORP) 19 December 1996 see the whole document ---	1-4,8
A	FOOTE S. ET AL.: "The human Y chromosome: overlapping DNA clones spanning the euchromatic region" SCIENCE, vol. 258, 2 October 1992, pages 60-66, XP002080220 see the whole document ---	1-4,8
P,X	LAHN B. AND PAGE D.: "Functional coherence of the human Y chromosome" SCIENCE, vol. 278, 24 October 1997, pages 675-680, XP002080221 see the whole document -----	1-4,8

INTERNATIONAL SEARCH REPORT

International application No.

PC/US 98/ 07115

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see continuation-sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-4,8 partially (subject 1. on continuation-sheet)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-4,8 partially

Isolated DNA which occurs on the non-recombining region of the human Y chromosome or the complement thereof, being testis-specific and optionally occurring in multiple copies on the Y chromosome.

Said DNA being the CDY gene, a characteristic portion, a probe or complement thereof, a DNA which hybridizes thereto under stringent conditions.

Said DNA having the SEQ ID NO:37,38 and coding for the amino acid of SEQ ID NO:39,40.

2. Claims: 1-4,8 partially

idem for BPY 1, SEQ ID NO:41,42

3. Claims: 1-4,8 partially

idem for BPY 2, SEQ ID NO:43,44

4. Claims: 1-4,8 partially

idem for XKRY, SEQ ID NO:45,46

5. Claims: 1-4,8 partially

idem for PTPRY, SEQ ID NO:47,48

6. Claims: 1-4,8 partially

idem for TTY 1, SEQ ID NO:49

7. Claims: 1-4,8 partially

idem for TTY 2, SEQ ID NO:50

8. Claims: 5-7,9 partially

Isolated DNA which occurs on the non-recombining region of the human Y chromosome or the complement thereof, not being testis-specific and having a homolog on the human X chromosome.

Said DNA being the DBY gene; a characteristic portion, a probe or complement thereof, a DNA which hybridizes thereto under stringent conditions. /

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Said DNA having the SEQ ID NO:17 and coding for the amino acid of SEQ ID NO:18.

9. Claims: 5-7,9 partially

idem for TPRY, SEQ ID NO:19,20,21,22,23,24

10. Claims: 5-7,9 partially

idem for TB4Y, SEQ ID NO:26,28

11. Claims: 5-7,9 partially

idem for EIF1AY, SEQ ID NO:30,32

12. Claims: 5-7,9 partially

idem for DFFRY, SEQ ID NO:34,36

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/07115

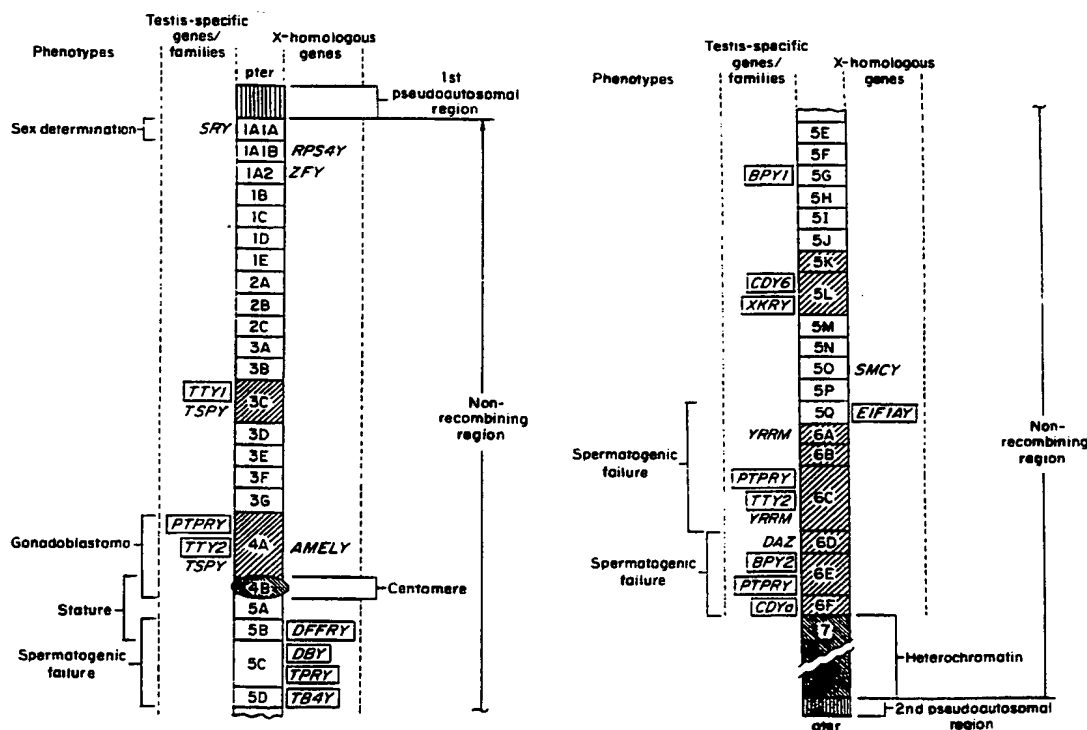
Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9200375 A	09-01-92	AU 670229 B AU 8093191 A CA 2085102 A EP 0536213 A	11-07-96 23-01-92 29-12-91 14-04-93
WO 9511300 A	27-04-95	AU 7947794 A	08-05-95
WO 9710267 A	20-03-97	AU 7156896 A CA 2238694 A EP 0859790 A	01-04-97 20-03-97 26-08-98
WO 9641007 A	19-12-96	US 5783390 A US 5776682 A AU 6159296 A CA 2221521 A EP 0832288 A US 5840549 A	21-07-98 07-07-98 30-12-96 19-12-96 01-04-98 24-11-98



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/12, 15/11		A3	(11) International Publication Number: WO 98/46747
			(43) International Publication Date: 22 October 1998 (22.10.98)
(21) International Application Number: PCT/US98/07115		(81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(22) International Filing Date: 10 April 1998 (10.04.98)			
(30) Priority Data: 60/041,877 11 April 1997 (11.04.97) US		Published With international search report.	
(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 60/041,877 (CIP) Filed on 11 April 1997 (11.04.97)		(88) Date of publication of the international search report: 4 March 1999 (04.03.99)	
(71) Applicant (for all designated States except US): WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US).			
(72) Inventors; and			
(75) Inventors/Applicants (for US only): LAHN, Bruce, T. [CN/US]; 863 Massachusetts Avenue #26, Cambridge, MA 02139 (US). PAGE, David, C. [US/US]; 3 Ivy Circle, Winchester, MA 01890 (US).			
(74) Agents: GRANAHAN, Patricia et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02173 (US).			

(54) Title: GENES IN THE NON-RECOMBINING REGION OF THE Y CHROMOSOME



(57) Abstract

Genes of the non-recombining region of the human Y chromosome, which fall into two classes: X-homologous DNA which is expressed in many organs and has functional X homologs and testis-specific DNA.

*(Referred to in PCT Gazette No. 13/1999, Section II)

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

GENES IN THE NON-RECOMBINING
REGION OF THE Y CHROMOSOME

GOVERNMENT SUPPORT

The invention described herein was made in whole or in
5 part with government support under Grant Number HG00257
awarded by the National Institutes of Health. The United
States Government has certain rights in the invention.

RELATED APPLICATIONS

This application claims the benefit of U.S.
10 Provisional Application No. 60/041,877, filed April 11,
1997, entitled "Genes in the Non-Recombining Region of the
Y Chromosome" by Bruce T. Lahn and David C. Page. The
entire teachings of the above referenced application is
expressly incorporated herein by reference.

15 BACKGROUND OF THE INVENTION

The human Y chromosome is distinguished from all other
nuclear chromosomes by four characteristics: the absence of
recombination, its presence in males only, its common
ancestry and persistent meiotic relationship with the X
20 chromosome, and the tendency of its genes to degenerate
during evolution (J. J. Bull, *Evolution of Sex Determining
Mechanisms* (Benjamin Cummings, Menlo Park, CA, 1983); J. A.
Graves, *Annu. Rev. Genet.* 30:233 (1996); B. Charlesworth,
Curr. Biol. 6:149 (1996); W. R. Rice, *BioScience*, 46, 331

-2-

(1996)). To be precise, these distinctive characteristics apply only to the non-recombining portion or region of the Y chromosome (NRY), which comprises 95% of the human Y chromosome. The remaining 5% of the chromosome is composed of two pseudoautosomal regions that maintain sequence identity with the X chromosome by meiotic recombination (H. J. Cooke et al., *Nature* 317:687 (1985); M. C. Simmler et al., *Nature* 317:692 (1985); D. Freije et al., *Science* 258:1784 (1992); G. A. Rappold, *Hum. Genet.* 92:315 (1993)).

Given the NRY's peculiar characteristics, one might expect its gene content to be idiosyncratic. Since discovery of the Y chromosome in 1923, its gene content has been the subject of speculation. By the middle of this century, while studies of human pedigrees had identified many traits exhibiting autosomal or X-linked inheritance, no convincing cases of Y-linked inheritance could be found (T. S. Painter, *J. Exp. Zool.* (1923); C. Stern, *Am. J. Hum. Genet.* 9:147 (1957)). As a result, consensus began to emerge that the Y chromosome carried few, if any, genes. In 1959, reports of XO females and XXY males established the existence of a sex-determining gene on the human Y chromosome (P. A. Jacobs et al. *Nature* 183:302 (1959); C. E. Ford et al., *Lancet*, i:711 (1959)), but this was perceived as a special case on a generally desolate chromosome. Opinions began to change only during the past decade, when eight NRY transcription units (or families of closely related transcription units) were identified, most during regionally focused, positional cloning experiments (D. C. Page et al., *Cell* 51:1091 (1987); A. H. Sinclair et al., *Nature* 346:240-244 (1990); J. Arnemann et al., *Genomics* 11: 108 (1991); E. C. Salido et al., *Am. J. Hum. Genet.* 50:303 (1992); E. M. Fisher et al., *Cell* 63:1205 (1990); K. Ma et al., *Cell* 75:1287 (1993); A. I. Agulnik et al., *Hum. Mol. Genet.* 3:879 (1994); R. Reijo et al., *Nat.*

-3-

Genet. 10:383 (1995)). It was not known if there were more genes in the NRY.

SUMMARY OF THE INVENTION

A systematic search of the non-recombining region of the human Y chromosome (NRY) has identified 12 novel genes or gene families. All 12 novel genes, and six of eight NRY genes or families previously isolated by less systematic means, fall into two classes. The first class of genes exists in one copy and is expressed in many organs; they have functional X homologs that escape X inactivation, as predicted for genes involved in Turner (XO) syndrome. The second class consists of Y-chromosomal gene families expressed specifically in testes, and may account for infertility among men with Y deletions.

The genes described herein, portions of the genes and DNA which hybridizes to genes or gene portions described are useful in diagnostic methods, such as a method to identify individuals in whom all or a portion of a gene or genes of the NRY is missing or altered. For example, Y chromosomal DNA from males with a known condition, such as infertility or reduced sperm count, can be assessed, using the gene(s) described herein, or characteristic portions thereof, to determine whether their DNA lacks some or all of the gene(s) described herein or contains an altered gene(s) (e.g., a gene in which there is a deletion, substitution, addition or mutation, compared to the sequences presented herein). Y chromosomal DNA (e.g., from a male with reduced sperm count or viability) can be assessed, using DNA described herein or DNA which hybridizes to DNA described herein, to determine whether the condition is associated with or caused by the occurrence of the gene or the gene alteration. For example, the presence or absence of all or a portion of a gene or genes shown to be necessary for fertility or

-4-

adequate sperm count can be assessed, using DNA which hybridizes to the gene or genes of interest to determine the basis for their infertility or reduced sperm count. In one embodiment, the occurrence of one or more Y-specific genes or a characteristic portion of one or more Y-specific genes is assessed in Y chromosomal DNA. In another embodiment, deletion or alteration of one of the testis-specific (Y-specific) genes described is assessed, such as by a hybridization method in which DNA which hybridizes to one of the Y-specific genes described herein or a characteristic portion thereof is used to assess a DNA sample obtained from a male who has a reduced sperm count. Lack of hybridization of the Y-specific DNA used to DNA in the sample indicates that the gene is not present in sample DNA or is present in an altered form which does not hybridize to Y-specific DNA of the present invention. In another embodiment, an X-homologous gene or genes present on the NRY can be used to determine whether the gene is present in an individual or if it occurs in an altered form in the individual. Using known methods, such as hybridization methods, X or Y chromosomal DNA from an individual can be assessed for the presence or absence of one or more of the X-homologous genes or a characteristic portion of one or more X-homologous genes. X or Y chromosomal DNA can also be assessed for the presence or absence of an altered form of one or more of the X-homologous genes described. In the present methods, DNA can be analyzed for the occurrence of Y-specific DNA, X-homologous genes or both. For example, a "battery" or group of DNA probes (sequences) can be used to analyze sample DNA; the probes can include Y-specific DNA probes (e.g., DNA which hybridizes to a Y-specific gene), X-homologous gene probes (e.g., DNA which hybridizes to an X-homologous gene) or both types of probes. DNA described herein is also useful as primers in an amplification

-5-

method, such as PCR, useful for identifying and amplifying Y-specific DNA or X-homologous genes in a sample (e.g., Y chromosomal DNA). Further, proteins or peptides encoded by the DNA described herein, such as proteins or peptides
5 encoded by an X-homologous gene or proteins or peptides encoded by testis-specific DNA (a testis-specific gene), can be assessed in samples. This can be carried out, for example, using antibodies which recognize proteins or peptides of the present invention (proteins or peptides
10 encoded by DNA described herein).

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a gene map of the non-recombining region of the Y chromosome.

Figure 2 shows the amino acid sequence alignments of
15 the chromodomain (SEQ ID NO.: 1-6) and putative catalytic domain (SEQ ID NO.: 7-12) of human CDY genes with their respective homologs. Amino acid identities are indicated by black shading and for each protein, the first and last amino acid residues are numbered (with respect to the
20 initiator methionine) and the total length of the protein is indicated. Chromodomain: SEQ ID NO.: 1, CDY (human); SEQ ID NO.: 2, HP1 (Drosophila); SEQ ID NO.: 3, Polycomb (Drosophila); SEQ ID NO.: 4, CHD1 (Drosophila); SEQ ID NO.:
5, Su(var) 3-9 (Drosophila); SEQ ID NO.: 6, PDD1 (Tetrahymena); SEQ ID NO.: 7; Covalent modification domain:
25 SEQ ID NO.: 8, CDY (human); SEQ ID NO.: 9, Enoyl-CoA Hydratase (Human); SEQ ID NO.: 10, 4-CBA-CoA dehalogenase (Arthrobacter); SEQ ID NO.: 11, Crotonase (C. acetobutylicum); SEQ ID NO.: 12, Naphthoate synthase (E.
30 coli).

Figures 3A and 3B are the nucleic acid sequence of DBX (long and short transcripts, SEQ ID NO: 13 and SEQ ID NO: 14, respectively) and the encoded amino acid sequences (SEQ ID NO: 15 and SEQ ID NO.: 16, respectively), DBY (SEQ ID

-6-

NO: 17) and the encoded amino acid sequence (SEQ ID NO: 18). Dots in the DBX DNA and protein sequences indicate that the nucleic acids or amino acid residues are the same as those represented for DBY; dashes indicate a missing
5 nucleic acid or amino acid residue.

Figures 4A and 4B present the nucleic acid sequences for three forms of TPRY (short, medium and long, SEQ ID NO: 19, SEQ ID NO: 20 and SEQ ID NO: 21, respectively) and the encoded amino acid sequences for the short, medium and long
10 forms (SEQ ID NO: 22, SEQ ID NO.: 23 and SEQ ID NO: 24, respectively).

Figure 5 presents the nucleic acid sequences of TB4X (SEQ ID NO: 25) and TB4Y (SEQ ID NO: 26) and the encoded amino acid sequences (SEQ ID NO: 27 and SEQ ID NO: 28,
15 respectively). Dots in the TB4X DNA and protein sequences indicate that the nucleic acids or amino acid residues are the same as those represented for TB4Y.

Figure 6 represents the nucleic acid sequences of EIF1AX (SEQ ID NO: 29) and EIF1AY (SEQ ID NO: 30) and the
20 encoded amino acid sequences (SEQ ID NO: 31 and SEQ ID NO: 32, respectively).

Figures 7A - 7D represent the nucleic acid sequences of DFFRX (SEQ ID NO: 33) and DFFRY (SEQ ID NO: 34) and the encoded amino acid sequences (SEQ ID NO: 35 and SEQ ID NO:
25 36, respectively).

Figure 8 represents the nucleic acid sequences of CDYa (SEQ ID NO: 37) and CDYb (SEQ ID NO: 38) and the encoded amino acid sequences (SEQ ID NO: 39 and SEQ ID NO: 40,
respectively).

30 Figure 9 represents the nucleic acid sequences of BPY1 (SEQ ID NO: 41) and the encoded amino acid sequence (SEQ ID NO: 42).

Figure 10 represents the nucleic acid sequence of BPY2 (SEQ ID NO: 43) and the encoded amino acid sequence (SEQ ID
35 NO: 44).

-7-

Figure 11 represents the nucleic acid sequences of XKRY (SEQ ID NO: 45) and the encoded amino acid sequence (SEQ ID NO: 46).

Figure 12 represents the nucleic acid sequences of
5 PTPRY (SEQ ID NO: 47) and the encoded amino acid sequence (SEQ ID NO: 48).

Figure 13 is the nucleic acid sequence of TTY1 (SEQ ID NO: 49).

Figure 14 is the nucleic acid sequence of TTY2 (SEQ ID
10 NO: 50).

Figure 15 shows the nucleic acid sequence of the human CDY Like (CDYL) gene, which is the human autosomal homolog of CDY, located on chromosome 6p and expressed ubiquitously.

15 Figure 16 shows the nucleic acid sequence of the mouse Cdyl (CDY like) gene, which is the mouse ortholog of human CDYL, located on chromosome 13 and expressed predominantly in the testis. A longer transcript of the gene is ubiquitously expressed.

20 Figures 17A - 17C show the nucleic acid sequences of human Variably Charged Protein family members VCP2r, VCP8r and VCP10r, which are expressed in the testis and highly polymorphic.

Figure 17A is the nucleic acid sequence of VCP2r.

25 Figure 17B is the nucleic acid sequence of VCP8r.

Figure 17C is the nucleic acid sequence of VCP10r.

DETAILED DESCRIPTION OF THE INVENTION

Y chromosome genes, classed as genes having X homologues and testis-specific (Y-specific) genes, are the
30 subject of the invention described herein, as are DNA which hybridize to (are complementary to) all or characteristic portions of the Y chromosome genes, the encoded products (e.g., proteins, peptides, glycoproteins), antibodies and methods of diagnosis or treatment in which the genes,

-8-

complementary DNA, encoded proteins or antibodies are used. As described herein, fragments that hybridized to Y chromosomal DNA were selected and then their nucleotide sequences determined. It was expected that these sequence fragments would represent a redundant sampling of a much smaller set of genes. Computer analysis revealed that 577 fragments corresponded to known Y genes, including seven of eight NRY genes and all eight pseudoautosomal genes previously reported. These findings suggested that the 2539 sequence fragments represented the great majority of all Y-chromosomal genes. After further analysis, both to eliminate human repetitive sequences and to assemble overlapping fragments into contigs, 912 novel and non-overlapping sequences were hybridized to Southern blots of human genomic DNAs. 308 sequences that detected at least one prominent male-specific fragment were judged likely to derive from the NRY, and for each work was carried out to isolate cDNA clones from a human testis library, as described in Example 1. Nucleotide sequencing of cDNA clones, and rescreening of libraries as necessary, yielded full-length cDNA sequences for ten novel NRY genes or families, and partial cDNA sequences for two additional ones (Table and Figures 1 - 14).

TABLE: 12 Novel Genes or Families in the NRY

Gene Symbol	Gene Name	Tissue Expression	Multi-copy on Y	X homolog	Escape x Inactivation
DBY	Dead Box Y	ubiquitous		DBX	yes
TB4Y	Thymosin β 4, Y isoform	ubiquitous		TB4X	yes
EIF1AY	Translation Initiation Factor 1A, Y isoform	ubiquitous		EIF1AX	yes
TPRY	TPR motif Y	ubiquitous		TPRX	yes
DFFRY	Drosophila Fat Facets Related Y	ubiquitous		DFFRX	yes
CDY	Chromodomain Y	testis	yes		
BPY1	Basic Protein Y 1	testis	yes		
BPY2	Basic Protein Y 2	testis	yes		
XKRY	XK Related Y	testis	yes		
PTPRY	Protein-Tyrosine Phosphatase Related Y	testis	yes		
TTY1	Testis Transcript Y 1	testis	yes		
TTY2	Testis Transcript Y 2	testis	yes		

-10-

All 12 novel genes were localized on the Y chromosome, as described in Example 2. Figure 1 is a gene map of NRY. As shown, the Y chromosome consists of a large non-recombining region (NRY; euchromatin plus heterochromatin) flanked by pseudoautosomal regions (pter, short arm telomere; qter, long arm telomere). The NRY is divided into 43 ordered intervals (1A1A through 7) which are defined by naturally occurring deletions (D. Vollrath, et al., *Science* 258:52 (1992)). Listed immediately above the Y chromosome in Figure 1 are nine NRY genes with functional X homologs; novel genes are boxed. Indicated immediately below the Y chromosome are 11 testis-specific genes or families, some with multiple locations. It is likely that some testis-specific families have members in additional deletion intervals; the locations indicated are representative, but are not necessarily exhaustive. At the bottom of Figure 1 are shown NRY regions implicated, by deletion mapping, in sex determination, germ cell tumorigenesis (gonadoblastoma), stature, and spermatogenic failure (K. Ma et al., *Cell* 75:1287 (1993); R. Reijo et al., *Nat. Genet.* 10:383 (1995); P. H. Vogt et al., *Hum. Mol. Genet.* 5:933 (1996); J. L. Pryor et al., *New England J. Med.* 336:534 (1997); K. Tsuchiya et al., *Am. J. Hum. Genet.* 57:1400 (1995); P. Salo et al., *Hum. Genet.* 95:283 (1995)). Euchromatic regions that are made up, at least partially, of Y-specific repeats are drawn in grey. *AMELY*, which appears to fall within such a repeat-containing region, is actually located in a sub-region of 4A that is not repetitive.

Expression of the 12 novel genes was assessed in diverse human tissues, by Northern blotting. -

Autoradiograms were produced by hybridizing ³²P-labeled cDNA probes to Northern blots of poly(A)⁺ RNAs (2 µg/lane) from human tissues (Clontech, Palo Alto, CA). Probes employed were cDNA clones, full-length (most genes) or

-11-

partial (*DBY*, nucleotides 1476-2319 of GenBank AF000985; *TPRY*, nucleotides 861-1768 of GenBank AF000996; *DDFRY*, nucleotides 8604-9878 of GenBank AF000986). Blots were hybridized at 65°C in Church's buffer (0.5 M Na₂PO₄ at pH7.5, with 7% SDS), and washed at 65°C in 1X SSC and 0.1% SDS. *DBY*, *TB4Y*, *EIF1AY* and *DDFRY* probes cross-hybridize to transcripts derived from their X homologs. For all five X-homologous genes (*DBY*, *TPRY*, *TB4Y*, *EIF1AY* and *DDFRY*), expression was tested and confirmed in three male tissues (brain, prostate and testis) by RT-PCR using Y-specific primers.

The novel genes encode an assortment of proteins and are dispersed throughout the euchromatic portions of the NRY. Nonetheless, all 12 genes fall into two discrete classes: 1) X-homologous genes and 2) testis-specific, Y-specific gene families (Table).

The X-homologous genes share the following characteristics: each has a homolog on the X chromosome encoding an extremely similar but nonidentical protein isoform, each is expressed in a wide range of human tissues (is not testis-specific), and each appears to exist in a single copy on the NRY. There are five novel representatives of this X-homologous class:

1. *DBY* encodes a novel "DEAD box" protein, perhaps an RNA helicase involved in translation initiation (P. Linder, et al., *Nature*, 337, 121 (1989); R.-Y. Chuang, P. L. Weaver, Z. Liu, T.-H. Chang, *Science*, 275, 1468 (1997)). The *DBY* protein is 91% identical to *DBX*, encoded by a homologous gene on the human X chromosome.
2. *TPRY* encodes a novel protein containing 10 tandem "TPR" motifs, a protein-protein interaction domain found in the products of the yeast *SSN6/CYC8*, *CDC16*, and *CDC23* genes, among others (R. S. Sikorski, M. S. Boguski, M. Goebel, P. Hieter, *Cell*, 60, 307 (1990); D. Tzamarias, K. Struhl, *Genes Dev*, 9, 821 (1995)). Differential splicing may

-12-

generate TPRY isoforms that differ at their carboxy termini. The amino terminal portion of the TPRY protein is 83% identical to TPRX, encoded by an homologous gene on the X chromosome.

5 3. *TB4Y* encodes a 44 amino acid protein that differs at only three residues from thymosin β_4 , which functions in actin sequestration (H. Gondo, et al., *J. Immunol.* 139:3840 (1987); D. Safer, M. Elzinga, V. T. Nachmias, *J Biol Chem*, 266, 4029 (1991)), and we found is located on the X. It is
10 proposed that the X-linked gene encoding thymosin β_4 be called *TB4X*.

4. *EIF1AY* encodes a Y-linked isoform of translation initiation factor 1A (eIF-1A) (T. E. Dever, et al., *J Biol Chem*, 269, 3212 (1994); J. W. Hershey, *Annu. Rev. Biochem.* 60, 717 (1991)), which we discovered is located on the X.
15 It is proposed that the X-linked gene encoding eIF-1A be called *EIF1AX*. The amino acid sequences of the X and Y-encoded proteins are 97% identical.

5. *DFFRY* encodes a Y-linked isoform of *DFFRX*, a recently
20 described X-linked protein. A Y-linked homolog was detected previously, but had been thought to be a pseudogene. The human *DFFRX* and *DFFRY* proteins, which are 91% identical, are homologous to the *Drosophila fat-facets* gene product, a deubiquinating enzyme required for eye
25 development and oogenesis (M. H. Jones, et al., *Hum Mol Genet* 5, 1695 (1996); J. A. Fischer-Vize, G. M. Rubin, R. Lehmann, *Development*, 116, 985 (1992); Y. Huang, R. T. Baker, J. A. Fischer-Vize, *Science*, 270, 1828 (1995)).

The second group of novel NRY genes, the testis-specific, Y-specific gene families, share a very different
30 set of characteristics: each appears to be expressed specifically in testes and each appears to exist in multiple copies on the NRY, as judged by i) the number and intensity of hybridizing fragments on genomic Southern
35 blots or ii) multiple map locations on the Y. We report

-13-

five novel testis-specific, Y-specific gene families with full-length cDNA sequences:

1. The *CDY* family encodes proteins with an amino-terminal "chromodomain," a chromatin binding motif (T. C. James, S. C. Elgin, *Mol Cell Biol*, 6, 3862 (1986); B. Tschiersch, et al., *EMBO J*, 13, 3822 (1994); R. Paro, D. S. Hogness, *Proc Natl Acad Sci U S A*, 88, 263 (1991); D. G. Stokes, K. D. Tartof, R. P. Perry, *Proc Natl Acad Sci U S A*, 93, 7137 (1996); M. T. Madireddi, et al., *Cell*, 87, 75 (1996)) (Figure 3). The carboxy-terminal half shows striking amino acid similarity, over a region of more than 200 residues, to nearly the full length of several enzymes, both prokaryotic and eukaryotic (M. Kanazawa, et al., *Enzyme Protein*, 47, 9 (1993); A. Schmitz, K. H. Gartemann, J. Fiedler, E. Grund, R. Eichenlaub, *Appl. Environ. Microbiol.* 258, 4068 (1992); Z. L. Boynton, G. N. Bennet, F. B. Rudolph, *J Bacteriol*, 178, 3015 (1996); V. Sharma, K. Suvarna, R. Meganathan, M. E. Hudspeth, *J Bacteriol*, 174, 5057 (1992); P. M. Palosaari, et al., *J Biol Chem*, 266, 10750 (1991)). The reactions catalyzed by these homologs are diverse, but in each case the substrate contains cofactor A (CoA) attached to a carbonyl group, and an alkoxide intermediate is formed. The unprecedented combination of a chromodomain and a putative CoA-substrate enzyme in a single polypeptide suggests that, in vivo, *CDY* proteins may catalyze covalent modification of DNA or chromosomal proteins, perhaps during spermatogenesis.
2. The *BPY1* genes encode a basic protein, 125 residues long, with little sequence similarity to known proteins. The encoded protein is rich in serine, lysine, arginine, and proline and has a pI of 9.4. Southern blotting studies revealed homologous sequences on the human X chromosome, but screening of cDNA libraries has failed to yield X-derived clones.

-14-

3. The *BPY2* genes encode a second basic protein, 106 residues in length, without obvious sequence similarity to *BPY1* or other known proteins. The pI of *BPY2* is 10.0.

4. The *XKRY* genes encode a protein with sequence
5 similarity to *XK*, a putative membrane transport protein defective in McLeod syndrome (M. Ho, et al., *Cell*, 77, 869 (1994)).

5. The *PTPRY* genes encode a protein with weak homology to a putative protein-tyrosine phosphatase (PTPase) in the
10 mouse (W. Hendriks, et al., *J Cell Biochem*, 59, 418 (1995)). Two additional families of testis-specific transcription units, referred to as *TTY1* and *TTY2*, have been identified. The sequences represented in Figures 14 and 15 are being assessed for open reading frames.

15 It appears that conventional single-copy genes, commonplace elsewhere in the genome, are quite uncommon in the NRY. Indeed, the two classes of NRY genes suggested by the systematic search described herein accommodate not only the 12 genes reported here, but also six of eight
20 previously identified NRY genes. *SRY*, a Y-specific gene that triggers the male pathway of sexual differentiation, is expressed in testes, and exists in only one copy in the NRY. *AMELY*, which has an X-linked homolog *AMELX*, is expressed only in the developing tooth bud. The X
25 inactivation status of *AMELX* is unknown.

Also described herein are five additional genes and their sequences (Figures 15, 16, 17A - 17C): human *CDY* Like (*CDYL*), which is the human homolog of *CDY*; it is on chromosome 6p and expressed ubiquitously; mouse *Cdyl* (*CDY*
30 like), which is the mouse ortholog of human *CDYL*; it is on chromosome 13 and expressed predominantly in testis and also has a longer transcript that is expressed ubiquitously; and human *VCP* (Variably Charged Protein) family, which is a family of genes on the X chromosome that
35 are homologous to *BPYI*, expressed in the testis and highly

-15-

polymorphic. Human CDY, human CDYL and mouse Cdyl have been shown to be histone acetyltransferases by *in vitro* assays. Human CDY is a candidate for the Azoospermia Factor (AZF) because it is within the AZFc region that is
5 commonly deleted in infertile men. Chemicals that block the enzymatic activity of any of these genes are candidate male contraceptives.

Inhibitors of the enzymatic activity of these genes, such as the human CDY gene, can be identified through an *in vitro* assay. For example, the protein encoded by one of
10 the genes (e.g., CDY-encoded protein) can be produced, such as by recombinant means (e.g., in bacterial cells containing a vector or plasmid which includes the gene to be expressed), and obtained. The effect of a candidate
15 inhibitor (drug) on the enzymatic activity of the protein can be assessed by combining the candidate inhibitor with the protein, a substrate of its enzymatic activity (e.g., histones) acetyl CoA (e.g., radiolabelled acetyl CoA) and other assay components (e.g., an appropriate physiological
20 solution or buffer), to produce a combination. The combination is maintained under conditions under which the enzymatic activity of the protein is maintained and appropriate for the protein to act upon/interact with its substrate (e.g., for the CDY gene to retain its histone
25 acetyltransferase activity). As a result, the substrate is acted upon by the protein if the candidate inhibitor does not inhibit the protein and the protein acts upon the substrate. If the substrate is not acted upon by the protein, this is an indication that the candidate inhibitor
30 is an inhibitor of the protein. For example, if a histone acetyltransferase, such as CDY-encoded protein is inhibited by a candidate inhibitor, its histone acetyltransferase activity will be blocked. If radiolabelled acetyl CoA is used, transfer of the radiolabelled acetyl group to the
35 enzyme substrate (histones) is inhibited (will not occur or

-16-

will occur to a lesser extent than occurs in the absence of the candidate inhibitor). Whether transfer occurs can be assessed by determining the location of radiolabelled acetyl groups from acetyl CoA. If the histone substrates are not radiolabelled or are radiolabelled to a lesser extent in the presence of a candidate inhibitor (than in its absence), the candidate inhibitor is an inhibitor of the protein. Inhibitors identified in this way can be further assessed in additional *in vitro* assays or in *in vivo* assays (e.g., in an appropriate animal model).

To interpret the observation that these X-homologous and multi-copy, testis-specific groups account for 18 of 20 known NRY genes or families, we postulate that the NRY's evolution was dominated by two strategies. The first strategy favors conservation of certain existing genes and the second favors the acquisition of a class of novel genes: 1) The X-homologous genes probably reflect the common ancestry of the X and Y chromosomes, and selective pressures to maintain comparable expression of genes in males and females. 2) The abundance of testis-specific families may have resulted from the NRY's selectively retaining and amplifying genes that enhance male reproductive fitness.

1) Dosage compensation and X-Y homology. Experts agree that the mammalian X and Y chromosomes evolved from autosomes, with nearly all ancestral gene functions deteriorating on the non-recombining portion of the emerging Y chromosome while being maintained on the nascent X chromosome (J. J. Bull, *Evolution of Sex Determining Mechanisms* (Benjamin Cummings, Menlo Park, CA, 1983); J. A. Graves, *Annu. Rev. Genet.* 30:233 (1996); B. Charlesworth, *Curr. Biol.* 6:149 (1996); W. R. Rice, *BioScience* 46:331 (1996)). Functional degeneration of the NRY would result in females having two, but males only one, copy of many genes, creating the need for a mechanism to equalize

-17-

X-linked gene expression in the sexes. In mammals, a predominant solution to this problem is provided by X inactivation, the transcriptional silencing of one X chromosome in females.

5 However, the findings on X-homologous NRY genes described herein, combined with previous studies, illustrate the importance in human evolution of an alternative solution: preservation of homologous genes on both the NRY and the X chromosome, with both male and
10 female cells expressing two copies of such genes. A critical prediction of this model is that, in female cells, the X homologs should escape X inactivation. This is the case for all widely expressed X-linked genes with known NRY homologs, including the X homologs of five novel NRY genes
15 reported here (E. M. Fisher, et al., *Cell* 63:1205 (1990); A. I. Agulnik et al., *Hum. Mol. Genet.* 3:879 (1994); M. H. Jones et al., *Hum. Mol. Genet.* 5:1695 (1996); J. A. Fischer-Vize et al., *Development* 116:985 (1992); Y. Huang et al., *Science* 270:1828 (1995); A. Schneider-Gädicke et
20 al., *Cell* 57:1247 (1989)). A second prediction of this model is that the human X and Y encoded proteins should be functionally interchangeable even though the nucleotide sequences of their corresponding genes are considerably diverged. Indeed, each of the eight known X-NRY gene pairs
25 encode closely related isoforms, with 83 to 97% amino acid identity throughout their lengths; functional interchangeability has been demonstrated in the one case tested to date (M. Watanabe et al., *Nat. Genet.* 4:268 (1993)).

30 Turner syndrome is classically associated with an XO sex chromosome constitution. In 1965, Ferguson-Smith postulated that the Turner phenotype might be due to inadequate expression of X-Y common genes that escape X inactivation (M. A. Ferguson-Smith, *J. Med. Genet.* 2:142
35 (1965)). These "Turner genes" have yet to be identified

-18-

with certainty. However, there now exists a substantial collection of X-homologous NRY genes (Figure 1) which can be assessed for genes which contribute to or are responsible for the Turner phenotype. The potential role of *RPS4Y* and *RPS4X* in Turner syndrome is controversial (E. M. Fisher et al., *Cell* 63:1205 (1990); W. Just et al., *Hum. Genet.* 89:240 (1992)). At least one Turner gene maps to the Xp-Yp pseudoautosomal region (T. Ogata et al., *J. Med. Genet.* 30:918 (1993)). Seven of the eight known X-NRY gene pairs appear to be ubiquitously expressed, and at least three encode housekeeping proteins: an essential ribosomal protein (*RPS4*), an essential translation initiation factor (*eIF-1A*), and a modulator of actin polymerization (thymosin β 4). Perhaps some features of the XO phenotype (e.g., poor fetal viability) reflect inadequate expression of such housekeeping functions.

2) Male fitness and Y-specific, testis-specific genes. As first appreciated by R.A. Fisher, animal genomes may contain genes or alleles that enhance male reproductive fitness but are inconsequential or detrimental with respect to female fitness (R. A. Fisher, *Biol. Rev.* 6:345 (1931)). As Fisher recognized, selective pressures would tend to favor the accumulation of such genes in male-specific regions of genomes. Of course, male reproductive fitness depends critically on sperm production, the central task of the adult testis. Since the NRY is the only male-specific portion of the mammalian genome, it should have a unique tendency to accumulate male-benefit genes during evolution.

These principles are illustrated by several gene families on the human NRY. *De novo* deletions of the *DAZ* gene cluster on the human Y chromosome are associated with severe spermatogenic defects (R. Reijo et al., *Nat. Genet.* 10:383 (1995)), and in *Drosophila* the *DAZ* homolog *boule* is required for spermatogenesis (C. G. Eberhart et al., *Nature* 381:783 (1996)). The *DAZ* gene cluster on the human Y

-19-

chromosome arose, during primate evolution, by transposition and amplification of an autosomal gene. Likewise, two other testis-specific NRY gene families —YRRM and TSPY — may also be the result of the Y chromosome's having acquired and amplified autosomal genes (R. Saxena et al., *Nat. Genet.* 14:292 (1996); M. L. Delbridge et al., *Nat. Genet.* 15:131 (1997)). It is possible that the selective advantage conferred by the NRY's retaining and amplifying male fertility factors (from throughout the genome) accounts for the multitude of testis-specific gene families there. This may have been the preeminent force in shaping the NRY's gene repertoire, as it appears that the great majority of NRY transcription units are members of such testis-specific families. In the NRY, each of the testis-specific gene families has multiple members, 20 to 40 copies in the case of TSPY (E. Manz et al., *Genomics* 17: 726 (1993)), and perhaps as many as 20 copies in the case of YRRM (K. Ma et al., *Cell* 75:1287 (1993)). All together, the various Y-specific gene families may include as many as several hundred genes or copies. Though it is not known how many of these are functional, it seems likely that Y-specific, testis-specific gene families comprise the great majority of NRY transcription units.

Recent genetic studies underscore the importance of the human Y chromosome in fertility. Many men with spermatogenic failure, but who are otherwise healthy, have deletions of portions of the NRY (K. Ma et al., *Cell* 75: 1287 (1993); R. Reijo et al., *Nat. Genet.* 10:383 (1995); P. H. Vogt et al., *Hum. Mol. Genet.* 5:933 (1996); J. L. Pryor et al., *New England J. Med.* 336:534 (1997)). These findings suggested the existence of NRY genes that play critical roles in male germ cell development but are not required elsewhere in the body. Previous deletion mapping studies have implicated four regions of the NRY in either

-20-

spermatogenic failure or germ cell tumorigenesis, and in each of these four regions we now report novel candidate genes expressed specifically, or most abundantly, in testes (Figure 1). As shown in Figure 1, the region implicated in gonadoblastoma, stature and spermatogenic failure all contain novel candidate genes. Two of the three regions implicated in spermatogenic failure each contain one or more novel testis-specific genes. The third region implicated in spermatogenic failure (intervals 5B-5D) contains two X-homologous genes, *DBY* and *EIF1AY*, with abundant, testis-specific transcripts in addition to higher-molecular-weight, ubiquitous transcripts.

While X-homologous and testis-specific genes are somewhat intermingled within the NRY, clustering is evident (Figure 1). The geographic distribution of the two classes correlates quite well with previously identified sequence domains within the euchromatic NRY (D. Vollrath et al., *Science* 258:52 (1992); S. Foote et al., *Science* 258:60 (1992)). Ten of the 11 known testis-specific families map to previously identified regions of Y-specific repetitive sequences. The only exception is *BPY1*, which cross-hybridizes to the X chromosome and maps to a previously recognized region of X homology. Indeed, one or more testis-specific gene families are found in nearly all known regions of euchromatic Y repeats (Figure 1). Ironically, it had been widely assumed that these regions consisted of "junk" DNA, partly on theoretical grounds (B. Charlesworth, *Science* 251:1030 (1991); E. Seaborn et al., *Cold Spring Harb. Symp. Quant. Biol.* 1:237 (1986)). To the contrary, the results presented here argue that these Y-specific repetitive regions contain the great majority of the NRY's transcription units (The only exception is *BPY1*, which cross-hybridizes to the X chromosome and maps to a previously recognized region of X homology). These regions may be the result of rampant gene amplification during

-21-

mammalian evolution. By contrast, none of the eight X-homologous genes map to the Y-repeat regions; all eight map to regions previously identified as consisting largely of single-copy (or in some cases X-homologous) sequences.

5 It is possible that, early in mammalian evolution, these regions of the NRY shared extensive sequence identity with the nascent X chromosome. The stage is now set for systematic evolutionary, biochemical and cell biological studies of the NRY, an idiosyncratic segment of the human
10 genome.

The present invention relates to isolated DNA and genes, present on (which occur on) the Y chromosome, whose sequences are provided herein, as well as characteristic portions of the DNA. It relates to additional nucleic
15 acid/nucleotide sequences which are not identical to the sequences presented herein but include substitutions or differences; DNA which includes substitutions or differences and encodes the same amino acid sequence as a DNA whose sequence is provided herein or includes
20 substitutions which do not alter the ability of a DNA probe or primer which hybridizes to DNA whose sequence is presented herein to hybridize to the DNA containing the substitutions or differences. It further relates to DNA which encodes a protein or peptide whose sequence is
25 presented herein. The present invention also includes the complements of the DNA sequences presented herein, DNA which hybridizes under stringent (high stringency) conditions to the DNA whose sequences are presented and to RNA transcripts. The invention further relates to encoded
30 proteins, peptides and other products (e.g., glycoproteins) and antibodies which are raised against or bind to proteins or peptides whose amino acid sequences are presented herein or are encoded by DNA whose sequences are provided. As used herein, the term isolated DNA which occurs on the non-
35 recombining region of the human Y chromosome refers to DNA

-22-

which has been obtained or removed from the human Y chromosome or DNA, produced by any means (e.g., recombinant techniques, synthetic methods), which has the sequence of such Y chromosome DNA. For example, isolated testis-specific DNA or isolated testis-specific DNA which occurs on the non-recombining region of the human Y chromosome is DNA which has been obtained or removed from the non-recombining region of the human Y chromosome or which has the sequence of such DNA and has been obtained or produced by any means.

Thus, this invention has application to several areas. It may be used diagnostically to identify males with reduced sperm count in whom a gene has been deleted or altered. It may also be used therapeutically in gene therapy treatments to remedy fertility disorders associated with deletion or alteration of a gene described. In one embodiment of a gene therapy method, a gene described herein, or a gene portion which encodes a functional protein, is introduced into a man whose sperm count is reduced and in whom the gene is expressed and the encoded protein replaces the protein normally produced or enhances the quantity produced. The present invention may also be useful in designing or identifying agents which function as a male contraceptive by inducing reduced sperm count. This invention also has application as a research tool, as the nucleotide sequences described herein have been localized to regions of the Y chromosome.

The present invention includes nucleotide sequences described herein, and their complements, which are useful as hybridization probes or primers for an amplification method, such as polymerase chain reaction (PCR), to show the presence, absence or disruption of the gene of the present invention. Probes and primers can have all or a portion of the nucleotide sequence (nucleic acid sequence) of a gene described herein or all or a portion of its

-23-

complement. For example, sequences shown in the Figures or Example 2 (SEQ ID NOS.: 1-84), as well as the complements thereof, can be used. The probes and primers can be any length, provided that they are of sufficient length and appropriate composition (appropriate nucleotide sequence) to hybridize to all or an identifying or characteristic portion of the gene described or to a disrupted form of the gene, and remain hybridized under the conditions use. Useful probes include, but are not limited to, nucleotide sequences which distinguish between a gene described herein and an altered form of that gene shown to be associated with reduced sperm count (azoospermia, oligospermia). Generally, the probe will be at least 7 nucleotides, while the upper limit is the length of the gene itself, e.g., up to about 40,000 nucleotides in length. Probes can be, for example, 10 to 14 nucleotides or longer (e.g., 20, 30, 50, 100, 250 nucleotides or any other useful length); the length of a specific probe will be determined by the assay in which it is used.

In one embodiment, the present invention is a method of diagnosing or aiding in the diagnosis of reduced sperm count associated with deletion or alteration of a gene described herein. Any man may be assessed with this method of diagnosis. In general, the man will have been at least preliminarily assessed, by another method, as having a reduced sperm count. By combining nucleic acid probes derived either from the isolated native sequence or cDNA sequence of the gene, or from appropriate primers, with the DNA from a sample to be assessed, under conditions suitable for hybridization of the probes with unaltered complementary nucleotide sequences in the sample but not with altered complementary nucleotide sequences, it can be determined whether the man possesses the intact gene. If the gene is unaltered, it may be concluded that the alteration of the gene is not responsible for the reduced

-24-

sperm count. This invention may also be used in a similar method wherein the hybridization conditions are such that the probes will hybridize only with altered DNA and not with unaltered sequences. The hybridized DNA can also be
5 isolated and sequenced to determine the precise nature of the alteration associated with the reduced sperm count. DNA assessed by the present method can be obtained from a variety of tissues and body fluids, such as blood or semen. In one embodiment, the above methods are carried out on DNA
10 obtained from a blood sample.

The invention also provides expression vectors containing a nucleotide (nucleic acid) sequence described herein, which is operably linked to at least one regulatory
15 nucleotide sequence is linked to a regulatory sequence in a manner which allows expression of the nucleotide sequence. The term "regulatory sequence" included promoters, enhancers, and other expression control elements (see, e.g., Goeddel, Gene Expression Technology: Methods in
20 Enzymology 185, Academic Press, San Diego, CA (1990)). It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the protein or peptide desired to be expressed. For instance, the peptides of the
25 present invention can be produced by ligating the cloned gene, or a portion thereof, into a vector suitable for expression in either prokaryotic cells, eukaryotic cells or both (see, for example, Broach, et al., Experimental Manipulation of Gene Expression, ed. M. Inouye (Academic
30 Press, 1983) p. 83; Molecular Cloning: A Laboratory Manual, 2nd Ed., ed. Sambrook et al. (Cold Spring Harbor Laboratory Press, 1989) Chapters 16 and 17).

Prokaryotic and eukaryotic host cells transfected by the described vectors are also provided by this invention.
35 For instance, cells which can be transfected with the

-25-

vectors of the present invention include, but are not limited to, bacterial cells such as *E. coli*, insect cells (baculovirus), yeast and mammalian cells, such as Chinese hamster ovary cells (CHO).

5 Thus, a nucleotide sequence described herein can be used to produce a recombinant form of the protein via microbial or eukaryotic cellular processes. Production of a recombinant form of the protein can be carried out using known techniques, such as by ligating the oligonucleotide
10 sequence into a DNA or RNA construct, such as an expression vector, and transforming or transfecting the construct into host cells, either eukaryotic (yeast, avian, insect or mammalian) or prokaryotic (bacterial cells). Similar procedures, or modifications thereof, can be employed to
15 prepare recombinant proteins according to the present invention by microbial means or tissue-culture technology.

The present invention also pertains to pharmaceutical compositions comprising the proteins and peptides described herein. For instance, the peptides or proteins of the
20 present invention can be formulated with a physiologically acceptable medium to prepare a pharmaceutical composition. The particular physiological medium may include, but is not limited to, water, buffered saline, polyols (e.g., glycerol, propylene glycol, liquid polyethylene glycol) and
25 dextrose solutions. The optimum concentration of the active ingredient(s) in the chosen medium can be determined empirically, according to procedures well known to medicinal chemists, and will depend on the ultimate pharmaceutical formulation desired. Methods of
30 introduction of exogenous polypeptides at the site of treatment include, but are not limited to, intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, oral and intranasal. Other suitable methods of introduction can also include rechargeable or biodegradable devices and slow release polymeric devices. The

-26-

pharmaceutical compositions of this invention can also be administered as part of a combinatorial therapy with other agents.

This invention also has utility in methods of treating disorders of reduced sperm count associated with deletion or alteration of a gene described herein. These genes may be used in a method of gene therapy, whereby the gene or a gene portion encoding a functional protein is inserted into cells in which the functional protein is expressed and from which it is generally secreted to remedy the deficiency caused by the defect in the native gene.

The present invention is also related to antibodies which bind a protein or peptide encoded by all or a portion of a gene of the present invention, as well as antibodies which bind the protein or peptide encoded by all or a portion of a disrupted form of the gene. For instance, polyclonal and monoclonal antibodies which bind to the described polypeptide or protein are within the scope of the invention. A mammal, such as a mouse, hamster or rabbit, can be immunized with an immunogenic form of the protein or peptide (an antigenic fragment of the protein or peptide which is capable of eliciting an antibody response). Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques are well known in the art. The protein or peptide can be administered in the presence of an adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibody.

Following immunization, anti-peptide antisera can be obtained, and if desired, polyclonal antibodies can be isolated from the serum. Monoclonal antibodies can be isolated from the serum. Monoclonal antibodies can also be produced by standard techniques which are well known in the

-27-

art (Koehler and Milstein, Nature 256: 495-497 (19775); Kozbar et al., Immunology Today 4: 72 (1983); and Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96 (1985)). Such antibodies are useful
5 as diagnostics for the intact or disrupted gene and also as research tools for identifying either the intact or disrupted gene.

The present invention is illustrated by the following examples, which are not intended to be limiting in any way.

10 EXAMPLE 1 ISOLATION OF CDNA CLONES FROM HUMAN TESTIS
LIBRARY

"cDNA selection" (M. Lovett et al., *Proc. Natl. Acad. Sci. USA* 88:9628 (1991)) was carried out using bulk cDNA prepared from human adult testes (Clontech, Palo Alto, CA)
15 and, as selector, a cosmid library prepared from flow-sorted Y chromosomes (Lawrence Livermore National Laboratory: LL0YNC03). A total of 3600 random cosmids, providing nearly five-fold coverage of the 30-Mb euchromatic region, were used to generate 150 pools of
20 selector DNA. Using each of the 150 selector pools, we carried out four successive rounds of cDNA selection, followed by two rounds of subtraction with human COT-1 DNA (Gibco BRL, Gaithersburg, MD) to remove highly repetitive sequences. A plasmid library was prepared from each of the
25 150 resulting pools of selected cDNA fragments, and 24 clones from each library were sequenced from one end. Of the 3600 sequences generated, about 600 were of poor technical quality and about 500 were found to derive from cloning vector or *E. coli* host, leaving 2539 sequences for
30 further analysis. Of the 2539 sequence fragments, 536 corresponded to previously reported NRY genes (487 to *TSPY*, 15 to *YRRM*, 14 to *RPS4Y*, 9 to *SMCY*, 5 to *DAZ*, 3 to *SRY*, 3 to *ZFY*) and 41 corresponded to previously reported pseudoautosomal genes (15 to *XE7*, 11 to *CSF2RA*, 4 to *IL3RA*,

-28-

3 to ASMT, 3 to IL9R, 2 to ANT3, 2 to MIC2, 1 to SYBL1). Electronic analysis of the roughly 2000 remaining sequences revealed that about 200 contained known repetitive elements, and these were not pursued. By electronically identifying redundancies and sequence overlaps, the remaining sequences were reduced to 1093 sequence contigs. Sequences representing these 1093 contigs were individually hybridized to dot-blotted yeast genomic DNAs of 60 YACs comprising most of the Y's euchromatic region (S. Foote et al., *Science* 258:60 (1992)). 181 sequences that hybridized to the great majority of the YACs were judged likely to contain highly repeated elements and were not pursued, leaving 912 sequences for further analysis. The 912 sequences were individually hybridized to Southern blots of R1-digested human 46,XX female and 49,XYYYY male (L. Sirota et al., *Clin. Genet.* 19:87 (1981)) genomic DNAs. Blots were hybridized at 65°C in Church's buffer (0.5 M Na₂PO₄ at pH7.5, with 7% SDS), and washed at 65°C in 1X SSC and 0.1% SDS, with 832 hybridizations yielding interpretable results. Many sequences appeared to contain highly repeated elements common to males and females, or failed to detect an unambiguously Y-specific restriction fragment, and these were not pursued. By contrast, 308 sequences hybridized to at least one prominent fragment present in 49,XYYYY but absent in 46,XX, suggesting that these sequences derived from the NRY. Each of these 308 sequences was individually used to screen, by hybridization, about 2 million plaques from a 1 phage library of human adult testis cDNA (Clontech, Palo Alto, CA).

EXAMPLE 2 LOCALIZATION OF 12 NOVEL GENES ON THE Y CHROMOSOME

Genes were localized on a previously reported NRY deletion map by testing with PCR for their presence or

-29-

absence in individuals carrying partial Y chromosomes (D. Vollrath et al., *Science* 258:52 (1992)). Most genes were localized to a single deletion interval. Some genes could not be unambiguously placed because copies exist in

5 multiple locations in the NRY. In such cases, genes were localized by PCR testing of YACs encompassing the NRY's euchromatic region (S. Foote et al., *Science* 258:60 (1992)). X homologs of Y genes were mapped onto the X by

10 PCR testing a panel of human/rodent somatic hybrid cell lines (Research Genetics, Huntsville, AL). All PCR assays consists of 30 cycles of the following conditions: 1 min denaturing at 94°C, 45 sec annealing at 60°C, and 45 sec extension at 72°C. TB4X primers were designed from an unreported intron. TPRX primers were designed from

15 unreported cDNA sequence. All other primers were designed from cDNA sequences as submitted to Genbank. PCR primers were as follows:

	GENE	LEFT PRIMER	RIGHT PRIMER
	DBY	CATTCGGTTTTTACCAGCCAG	CAGTGACTCGAGGTTCAATG
20		(SEQ ID NO.: 51)	(SEQ ID NO.: 52)
	TPRY	GCATCATAATATGGATCTAGTAGG	GGAGATACTGAATAGCATAGC
		(SEQ ID NO.: 53)	(SEQ ID NO.: 54)
	TB4Y	CAAAGACCTGCTGACAATGG	CTCCGCTAAGTCTTTTACC
		(SEQ ID NO.: 55)	(SEQ ID NO.: 56)
25	EIF1AY	CTCTGTAGCCAGCCTCTTC	GACTCCTTTCTGGCGGTAC
		(SEQ ID NO.: 570)	(SEQ ID NO.: 58)
	DFFRY	GAGCCCATCTTTGTCAGTTTAC	CTGCCAATTTTCCACATCAACC
		(SEQ ID NO.: 59)	(SEQ ID NO.: 60)
	CDY	GGCTCAAAATCCACTGACG	CAAGCGATATCTCACCACC
30		(SEQ ID NO.: 61)	(SEQ ID NO.: 62)
	BPY1	CTCCCTGAGCAGCAACTAAG	GTCATCAACATGGGAAGCAC
		(SEQ ID NO.: 63)	(SEQ ID NO.: 64)
	BPY2	CCAGGACCATGTGATATGG	CTAATTCCCTCTTTACGCATGACC
		(SEQ ID NO.: 65)	(SEQ ID NO.: 66)

-30-

<i>XKRY</i>	CACTCATGGAGAAGGGTAGG (SEQ ID NO.: 67)	GTCACACTCAGCCTCTTTAC (SEQ ID NO.: 68)
<i>PTPRY</i>	GAGCACACCACACCAGAAAC (SEQ ID NO.: 69)	CTCAGACTGACCTCGGACTG (SEQ ID NO.: 70)
5 <i>TTY1</i>	CTCTGGGAATCAAATTCGAGG (SEQ ID NO.: 71)	GTCTTTCAGCCAATCCAAGG (SEQ ID NO.: 72)
<i>TTY2</i>	GACAACTCTGACAGCCAGG (SEQ ID NO.: 73)	GTCAGAACTCCCAAACAGG (SEQ ID NO.: 74)
<i>DBX</i>	CTACATGCAGATGACATGGTG (SEQ ID NO.: 75)	GGCCAAGGTGCATAGGTG (SEQ ID NO.: 76)
10 <i>TPRX</i>	CATGTTCCCTGTAGCACATC (SEQ ID NO.: 77)	CGTTTCCATTACTTCCATTTCCTG (SEQ ID NO.: 78)
<i>TB4X</i>	CCCGCCCTTTCATCATCC (SEQ ID NO.: 79)	GCTCCCCAAAGTAGCCTTC (SEQ ID NO.: 80)
15 <i>EIF1AX</i>	CACGAGGCGCCATTTGCTG (SEQ ID NO.: 81)	CTGGAGGCCAGGCAACGTG (SEQ ID NO.: 82)
<i>DFFRX</i>	CCTCCACCTGAAGATGCC (SEQ ID NO.: 83)	CTGAGATCCAGGTGAATGG (SEQ ID NO.: 84)

EQUIVALENTS

20 Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

-31-

CLAIMS

We claim:

1. Isolated testis-specific DNA which occurs on the non-recombining region of the human Y chromosome or the complement thereof.
5
2. The isolated testis-specific DNA of Claim 1 which occurs in multiple copies on the non-recombining region of the human Y chromosome or the complement thereof.
3. The isolated testis-specific DNA of Claim 2 selected
10 from the group consisting of:
 - (a) a CDY gene or a characteristic portion thereof;
 - (b) a BPY 1 gene or a characteristic portion thereof;
 - (c) a BPY 2 gene or a characteristic portion thereof;
 - (d) an XKRY gene or a characteristic portion thereof;
 - 15 (e) a PTPRY gene or a characteristic portion thereof;
 - (f) TTY1 DNA; or a characteristic portion thereof;
 - (g) TTY 2 DNA; or a characteristic portion thereof;
 - (h) a complement of (a);
 - (i) a complement of (b);
 - 20 (j) a complement of (c);
 - (k) a complement of (d);
 - (l) a complement of (e);
 - (m) a complement of (f);
 - (n) a complement of (g);
 - 25 (o) DNA encoding the amino acid sequence of SEQ ID No.: 39;.
 - (p) DNA encoding the amino acid sequence of SEQ ID No.: 40;
 - (q) DNA encoding the amino acid sequence of SEQ ID
30 No.: 42;
 - (r) DNA encoding the amino acid sequence of SEQ ID No.: 44;

-32-

- (s) DNA encoding the amino acid sequence of SEQ ID No.: 46;
 - (t) DNA encoding the amino acid sequence of SEQ ID No.: 48; and
 - 5 (u) DNA which hybridizes to a DNA of any one of (a) through (t) under stringent conditions.
4. Isolated testis specific DNA selected from the group consisting of:
- (a) DNA of SEQ ID No.: 37;
 - 10 (b) DNA of SEQ ID No.: 38;
 - (c) DNA of SEQ ID No.: 41;
 - (d) DNA of SEQ ID No.: 43;
 - (e) DNA of SEQ ID No.: 45;
 - (f) DNA of SEQ ID No.: 47;
 - 15 (g) DNA of SEQ ID No.: 49;
 - (h) DNA of SEQ ID No.: 50;
 - (i) DNA encoding the amino acid sequence of SEQ ID No.39;
 - (j) DNA encoding the amino acid sequence of SEQ ID No.40;
 - 20 (k) DNA encoding the amino acid sequence of SEQ ID No.42;
 - (l) DNA encoding the amino acid sequence of SEQ ID No.44;
 - 25 (m) DNA encoding the amino acid sequence of SEQ ID No.46;
 - (n) DNA encoding the amino acid sequence of SEQ ID No.48;
 - (o) a complement of a DNA of any one of (a) through (n); and
 - 30 (p) DNA which hybridizes to a DNA of any one of (a) through (o) under stringent conditions.

-33-

5. Isolated X-homologous DNA which occurs on the non-recombining region of the human Y chromosome, is not testis-specific and has a homolog on the human X chromosome.
- 5
6. The isolated DNA of Claim 5 selected from the group consisting of:
- (a) a DBY gene or a characteristic portion thereof;
 - (b) a TPRY gene or a characteristic portion thereof;
 - 10 (c) a TB4Y gene or a characteristic portion thereof;
 - (d) an EIF1AY gene or a characteristic portion thereof;
 - (e) a DFFRY gene or a characteristic portion thereof;
 - 15 (f) a complement of (a);
 - (g) a complement of (b);
 - (h) a complement of (c);
 - (i) a complement of (d);
 - (j) a complement of (e);
 - 20 (k) a complement of (f);
 - (l) DNA encoding the amino acid sequence of SEQ ID No.: 18;
 - (m) DNA encoding the amino acid sequence of SEQ ID No.: 22;
 - 25 (n) DNA encoding the amino acid sequence of SEQ ID No.: 23
 - (o) DNA encoding the amino acid sequence of SEQ ID No.: 24;
 - (p) DNA encoding the amino acid sequence of SEQ ID No.: 28;
 - 30 (q) DNA encoding the amino acid sequence of SEQ ID No.: 32;
 - (r) DNA encoding the amino acid sequence of SEQ ID No.: 36; and;

-34-

- (s) DNA which hybridizes to a DNA of any one of (a) through (r) under stringent conditions.

7. Isolated X-homologous human DNA selected from the group consisting of:

- 5 (a) DNA of SEQ ID No.: 17 or a characteristic portion thereof;
- (b) DNA of SEQ ID No.: 19 or a characteristic portion thereof;
- 10 (c) DNA of SEQ ID No.: 20 or a characteristic portion thereof;
- (d) DNA of SEQ ID No.: 21 or a characteristic portion thereof;
- (e) DNA of SEQ ID No.: 26 or a characteristic portion thereof;
- 15 (f) DNA of SEQ ID No.: 30 or a characteristic portion thereof;
- (g) DNA of SEQ ID No.: 34 or a characteristic portion thereof;
- (h) DNA encoding the amino acid sequence of SEQ ID
- 20 No.: 18;
- (i) DNA encoding the amino acid sequence of SEQ ID No.: 22;
- (j) DNA encoding the amino acid sequence of SEQ ID No.: 23;
- 25 (k) DNA encoding the amino acid sequence of SEQ ID No.: 24;
- (l) DNA encoding the amino acid sequence of SEQ ID No.: 28;
- (m) DNA encoding the amino acid sequence of SEQ ID
- 30 No.: 32;
- (n) DNA encoding the amino acid sequence of SEQ ID No.: 36;
- (o) a complement of a DNA of any one of (a) through (n); and

-35-

- (p) DNA which hybridizes to a DNA any one of (a) through (o) under stringent conditions.
8. A DNA probe comprising all or a characteristic portion of DNA of Claim 4.
- 5 9. A DNA probe comprising all or a characteristic portion of DNA of Claim 7.

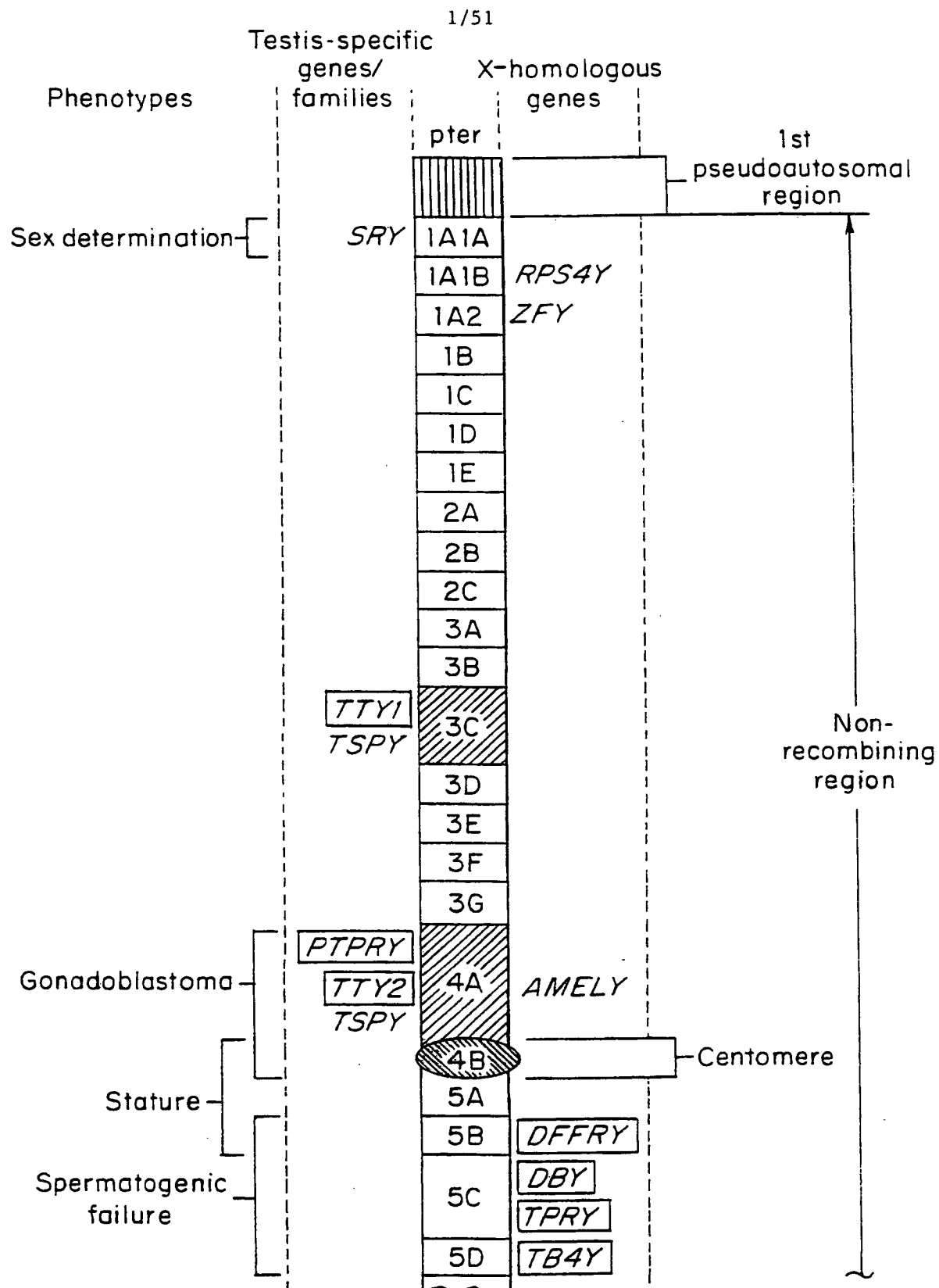


FIG. 1A

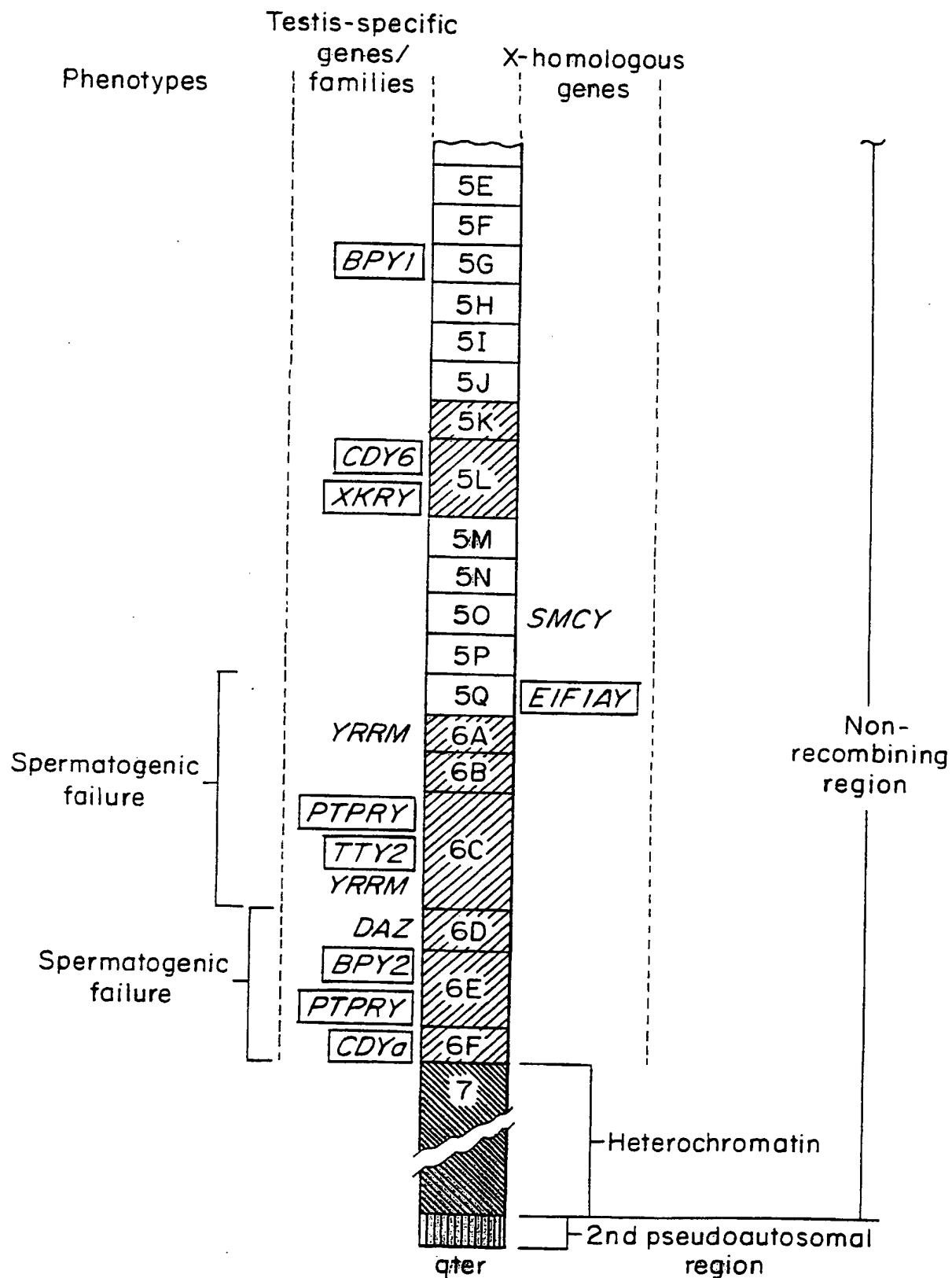


FIG. 1B

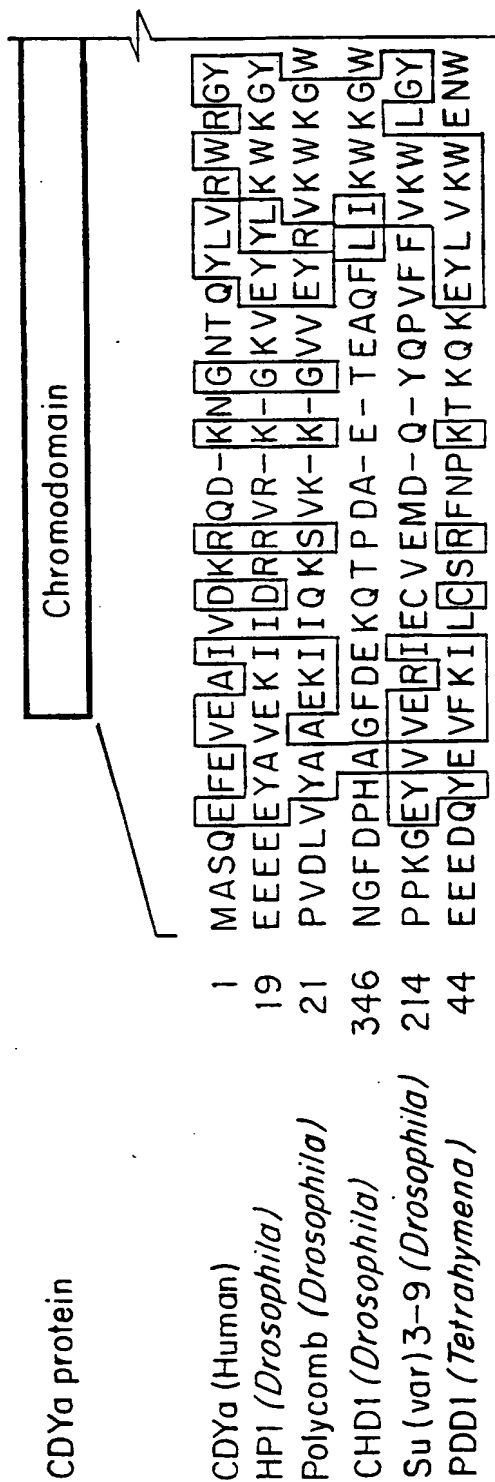


FIG. 2A

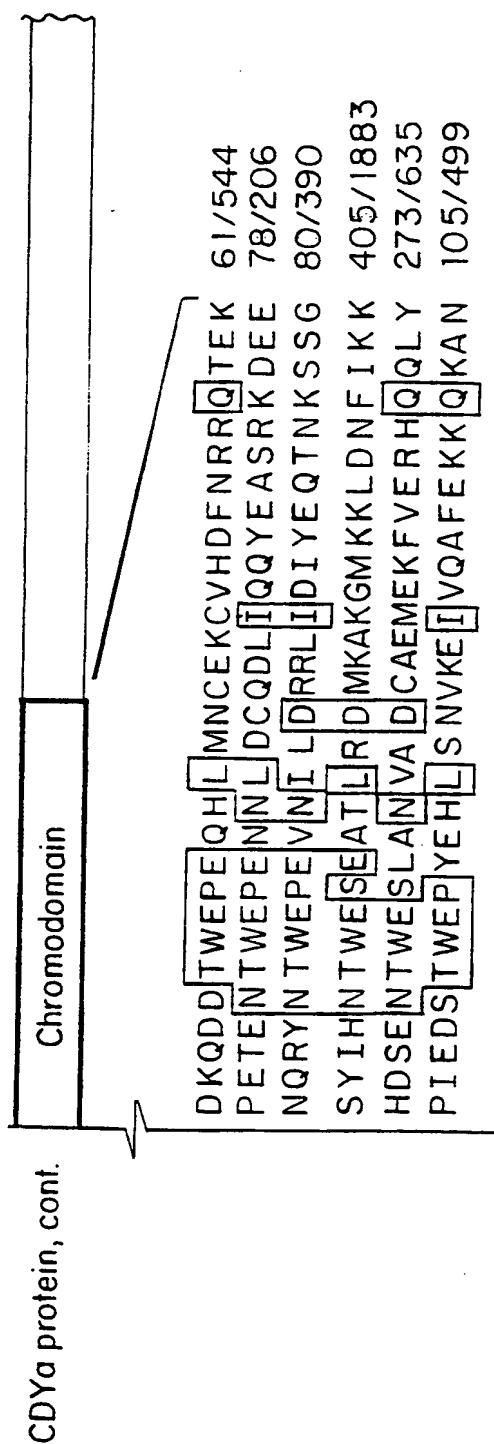


FIG. 2B

5/51

CDYa protein, cont		Covalent Modification Domain	
CDYa (Human)	246	SVPRVKGQQRNITDDSRDQPF	IKKMHFTIRLT
Enoyl-CoA Hydratase (Human)	1	MAALRVLLSCARGPLRPPVRC	PAWRPFASG
4-CBA-CoA dehalogenase (<i>Arthrobacter</i>)	1		
Camitine Racemase (<i>E. coli</i>)	1	MKRQGTTLPANNHALKQYAFF	AGMLSSLKKQK
Crotonase (<i>C. acetobutylicum</i>)	1		
Naphtholate Synthase (<i>E. coli</i>)	1	MIYPDEAML	YAPVEWHD
		Linker	
	349	YFVKHLRNNRNTASLEMVDT	IKNFVNTFIQFK
	102	EM-----QNL	SFQDCYSSKFLKHGHL
	72	EVPMGPASEIQSHFRLKAL	YYHVIHML
	105	AA-----AE	GEAPDADFGPGGFAGL
	70	EM-----KEMNT	IEGRKFGILGNKVFRRL
	90	VR	GDYGGYKDDSGVHHLNVLD
			FQRQIRTCP
	454	REACAKGLV	SQVFLTGFTQ
	201	QDAKQAGLV	SKICP
	177	DEAVEWGVVNR	VFSEADFQSRVGE
	205	EALRWGLVNR	VVSQAELMDN
	171	DEALRI	GLVNIKVV
	193	KQALDMGLVNT	VVPLADLEKE
			TVRWCREMLQN

FIG. 2C

CDYa protein, cont.

Covalent Modification Domain

ESASTYRDI VVKKE DGF TQ I V L S T R S T E K N A L N T E V I K E I
 ANFEYIIAEKRGKNNTVGL IQL - M R P K A L N A L C D G L I D E L
 MSSNSDHHISVEHT D G V A T I R F - T R P S K H N A S G Q L L L E T
 WRKGMSESLHLTRNGSILE I T L - D R P K A - N A I D A K I S F E M
 MELNNVILE K E G K V A V V I I N R P K A L N A I N S D I L K E M
 C S E G F E D I R Y E K S T D G I A K I T I - N R P Q V R N A F R P L I V K E M

K P I V V S V N G P A I G L G A S I L P L C D L V W A N E K A W F Q T P Y T T F
 K P V I A A V N G Y P F G G G C E L A M M C D I I Y A G E K A Q F A Q P E I L I
 K P T L A A I N G P A V G G G L G M S L A C D I A V C T D R A T F L P A W M S I
 K P V I A A V N G Y A F G G G F E L A I A A D F I V C A D N A S F A L P E A K L
 K P V I A A V N G F A I L G G G C E I A M S C D I R I A S S N A R F G Q P E V G L
 K P V V A M V A G Y S I G G G H V L H M M C D L T I A A D N A I F G Q T G P K V

N P I V L E E C K A L V R C N I K L E L E Q A N E R E C E V L R K I W S S A R G
 S K I V V A M A K E S V N A A F E M T L T E G S K L E K K L F Y S T F A T D D R
 P T H L Q G L V K N R I Q E G S S E T L E S C T E H E V Q N V I A S V G H P H F
 A P L A I A A L K E I Y R T T S E M P V E E A Y R Y I R S G V L K H Y P S V L H
 A P V A V K L S K Q A I N R G M Q C D I D T A L A F E S E A F G E C F S T E D Q
 S P M A L R C L K A A L N A D C D G Q A G L Q E L A G N A T M L F Y M T E E G Q

FIG. 2D

7/51

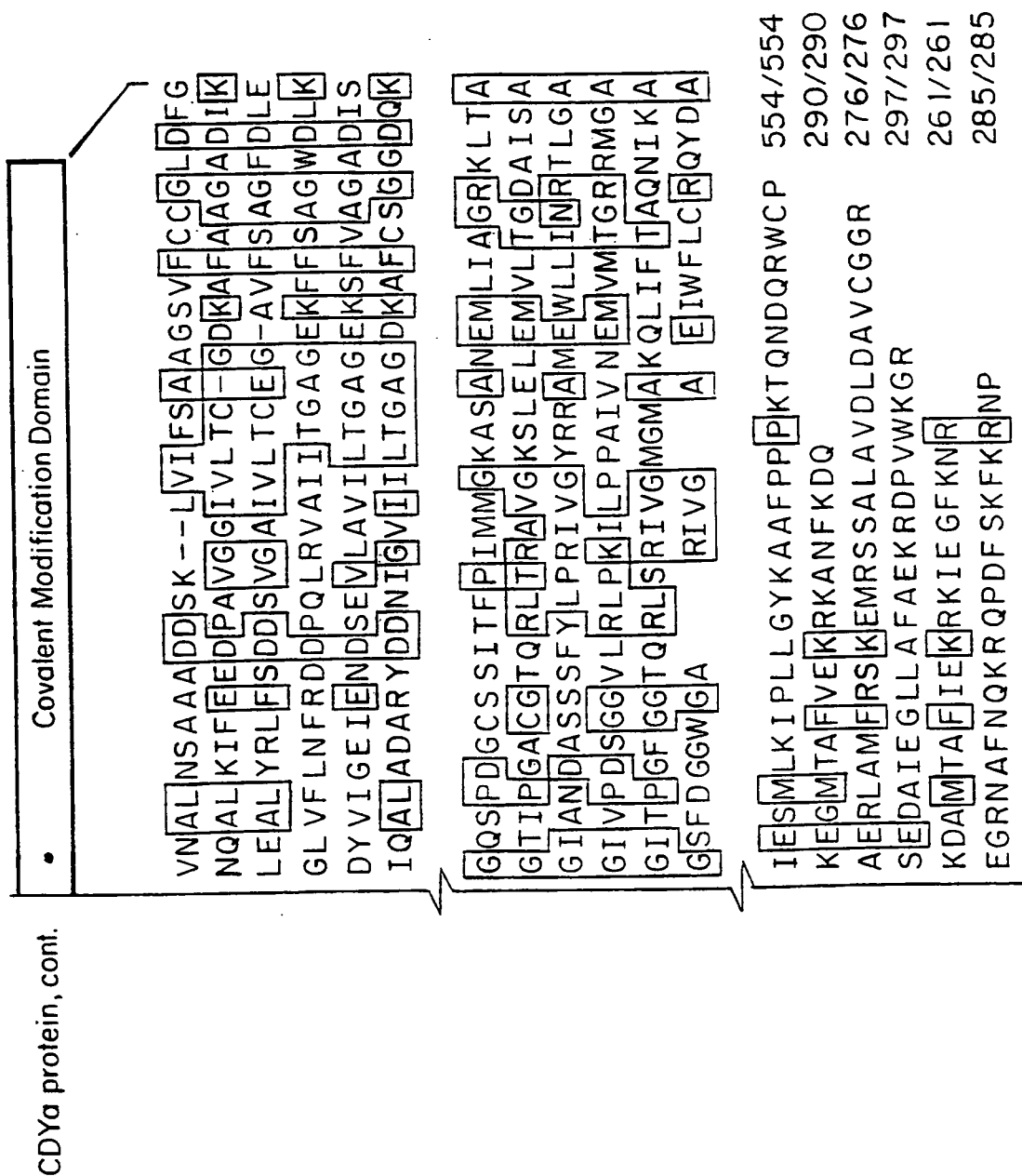


FIG. 2E

[illegible]

FIG. 3A

[illegible]

FIG. 3B

11/51

```

598      . . . . . C . . . . . A . . . . . S . . . . . C . . . . . H . . . . . T
1798  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
1792  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
600  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
      . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
628  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
1888  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
1882  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
630  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
      . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
658  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
1978  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
1972  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
660  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
      . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
2067  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
2060  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
      . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
2186  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
2155  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
2142  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
      . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
2245  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
2240  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
2232  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
2239  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
      . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
2335  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
2330  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
2322  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
      . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
2425  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
2420  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .
2412  DBX  . . . . . C . . . . . A . . . . . S . . . . . C . . . . . C . . . . .

```

FIG. 3D

12/51

[illegible]

FIG. 3E

[illegible]

FIG. 3F

14/51

TPRY
short, medium and long transcripts

```

-1005      gctcatcgttttgttg
-990      tctctttagacttgggtcagtgagaggaataaggcaaaagaaaccagcctatggagtgacaaggccttagggccaaagtcttgagggtga
-810      aggttttagggcctgcgcagcttccctgccatgccccgcaaggtctgcattcgcaaggcttgacagtgaggcctcattacgggactct
-720      cctaaagtcccatgggtgctctcttttcgcatttgcgccccgtgggtgatgcccgatgccgcttcccatcgctcttcccccttcaagcg
-630      tatcgcaactgcataaaacacccagcacagacactcatttctatctaatgcatttaactagcacacacctacaggtgttccatccccag
-540      agactaccccttttccatagacgtgacctcaacacccagcgggtcagaatcagtcagcctctgtcatgttccctaggctcttggegaac
-450      tggctggcggggtccagcagcctaggagtagagtgagcaatgcctgacgtgaagtcacaagaagatcacgtgagacgaatcagtcgcct
-360      agattggctacaactaaagtgttgagcgaggaggtcgcgcgctgcgtggggttcgccccgtgacacaaattacaaactttgtgctggtg
-270      ctggcaaaagtttgtgattttaagaaaattctgctgctccagcactgcgagcttctgccttccctgtagtttccagatgtgatccag
-180      gtacggaggtccgctgcccgtgcttcggtagcttaagtcttgcctcagcttttcccttgagccgctgagggcgatataaaattggc
-90      gtcacagtcctcaagcagcgattggaaggcgtcttcttcaactactcgttaaggttgggtatcgtcgtgggacttggaaatttgtgttcc

1      ATGAAATCTTGGCCAGTGTGGCTCACATACCGCGCGCTGTGTGGCTTCGGTGTGATGAGGCAAGAANAATGGCGGAAGGAAAGCGAGCCCGGAG
1      M K S C A V S L T T A A V A F G D E A K K M A E G K A S R E
91      AGTGAAGAGGAGTCTGTAGCTTGACAGATCGAGGAAAGGAGGCGCTTGCTTGGCATGGACACAGCCGTCTCTTCGGGTTCGTGAGGCTTCAT
31      S E E S V S L T V E R E A L G G M D S R L F G F V R L H
181      GAAGATGGCGCCAGAACGACCCCTACTTAGGCAAGGCTGTTCGCTGCTACGAATCTTTTAATCTTTAAAGCTGAAGGAAAGTGGAGTCT
61      E D G A R T K T L L G K A V R C Y E S I I L K A E G K V E S

```

FIG. 4A

[illegible]

FIG. 4B

17/51

Medium Short
1051 T K L P A F A R V V S A G N L L T H V G H T I L G M N T V Q
3151 ACTAACTTCCTGCTTTTGGCGGTGTCAGCAGGAAATCTTCAACCCATGTTGGGCATACCATGACAGTACAA
3151 GCTTGTCTTGAACCTCCCTGACCTCAGGTGTCCTGCTTCCCAAGTCTGGGATTTACAGTGTGAGCCACCATGCCCGTAA
1051 A C L E L L T S G G L L A S A S Q S A C I T G V S H H A R 1079

Medium Short
1081 L Y M K V P G S R T P G H Q E N N N F C S V N I N I G P G D
3241 CTGTATGAAAGTTCAGGGAGTCGGACACAGGTCAACCAAGAAATACAACTTCTGCTCTGTTAACATAATATTTGGTCCAGGAGAT
3241 acttttaaaatgttaagcaaaattacagatagtataaaacacacacattgtctaatggagaaataaaagtctcctacttttacatctaaaaa 3330

Medium
1111 C E W F V V P E D Y W G V L N D F C E K N N L N F L M S S W
3331 TGTGAATGGTTTGTGTACCTGAAGATTATTGGGTGTTCTGAATGACTTCTGTGAAATAATAATTTGAATTTTAAATGAGTCTTGG

Medium
1141 W P N L E D L Y E A N V P V Y R F I Q R P G D L V W I N A G
3421 TGGCCCAACCTGAAGATCTTTATGANGCAATGTCCTGTGTATAGATTTATTCAGCGACCTGGAGATTGGTCTGGATAAATGCAGGC

Medium
1171 T V H W V Q T V G W C N N I A W N V G P L T A C Q Y K L A V
3511 ACTGTGCAATGGGTTCAAACTGTTGGCTGGTGCAATAACATTGCCTGGAATGTTGGTCCACTTACAGCCTGCCAGTATAAATTTGGCAGTG

Medium
1201 E R Y E W N K L K S V K S P V P M V H L S W N M A R N I K V
3601 GAACGGTATGAATGGAAACAAATTGAAAAGTGTGAAGTCACCCAGTACCCTGGTGCATCTTTCTCTGGAATATGGCACGAAATATCAAAGTC

FIG. 4D

18/51

Medium Long	1231 3691 3723 3741	<p>S D P K L F E M I K * 1240 TCAGATCCAAAGCTTTTGAATGATTAAGTAAgtgccttctgaaactgctgcagtttctcttcttgggggtattggtagccattcagttatt Y C L L K I L K Q Y Q T L R E A L V A A TTGTCTTTTGAATAATCTGAGCAATATCAGACATTTGACAGAGCTCTTGTGTCAGCA</p>
Medium Long	3781 3781 1261	<p>tttttcaaaagaattctgttgacattaaatgatatacagcagtcagaaagtcttggcaaaatgtaataagatgtaaaataatcttatatt GGAAAAGAGGTTATATGCGCATGGCGGACAAATGATGAAACAGCTCATTACTGTAGCATTGTGAGGTGAGGTTTAAATCTGCTTTT G K E V I W H G R T N D E P A H Y C S I C E V E V F N L L F</p>
Medium Long	3871 3871 1291	<p>cataagtgttataaaatctcataagattaaatattgccttcccttaaaaaa 3926 GTCACATAAGAAAGCAATACTCAAAATACCTACATAGTACATTCCTTGTGACGAAACAAAGCAAAAGTTTGGAAAATTTTGTG V T N E S N T Q K T Y I V H C H D C A R K T S K S L E N F V</p>
Long	3961 1321	<p>GTGCTCGAACAGTACAAAATGGAGGACCTAATCCAAGTTTATGATCAATTTACACTAGCTCTTTTCATTTATCATCTCATCTTGATatagc V L E Q Y K M E D L I Q V Y D Q F T L A L S L S S S * 1347</p>

FIG. 4E

Long	4051	tccatgaatatataatgagattttctctgctcttcaggaaatttctgcaccactgggtttgtgagctgttttcataaaaactgttgactaaaa
Long	4141	gctatgtctatgcaaccttccaagaatagtagtcaagcaactggacacacagtgctgacctgtgcttcaggacttaacatgctgataccagct
Long	4231	gtacttcagaaaaataatataatcataatgtttgtgtacgtatgacaaaactgtcaaaagtgcacagaaatactgatttgaagatagcctt
Long	4321	ttttatgtttctctatttctgggctgatgaatttaattcatttgtattttaacctggagaatttcccttagttaaaaacactttccta
Long	4411	gctggtcatttcttcataagataagataagcaatttaaatctctcctcgatcagcttttaaaaatgtgtactattatctgaggaagtgtttttac
Long	4501	tgccttatgttttctgtgttttgaggccatgatgtacatttgggttccaaaataaatttttttaaatatttaataagcccatatacaaaa
Long	4591	gataatggattgcacatagacaaagaaaataaaacttcagatttgtgattttctctaaacttgatacacagatttacactatttataaata
Long	4681	cgtatttatgtcctgaaaaatatttgtgaatggaatgttgttttttccagacgtaaactgcacataaatactactaaggagttctgtagtttta
Long	4771	aacactactcctattacatttttatatgtgtagataaaactgcttagtattatcacagaaatttttataaaaattgttaaatgttttaaaggg
Long	4861	tttcccaatgtttgagtttaaaaagactttctgaaaaaatccacttttgttccatttccaaacctaaatgattatgtatttttatatgt
Long	4951	gtgtgtatgtgtacacacatgtataatataacagaaacctcgatataatgtgtatagattttaaagtgtttatttttacatctatgg
Long	5041	tagtttttgagggtgcctattataaagtattacggaagtgtgctgtttttaaagtaaatgtcttttagtgtgatttattaaagtgttagtca
Long	5131	ccatagtgatagcccatataaataatgtctggaataatgtatttataacagtagaaaacatatagtcagtgaaagtaaatatttttaaaggaa
Long	5221	acattatataagatttgataaaatgttgtttataatgaagagtttcttatggaaaagagattccagaatgataacctcttttagagaaacaaat
Long	5311	aagtgaacttatttttttaaagctagatgactttgaaatgctataactgtcctgcttgtaacaacatgggttggggtgaaggggaggaagaaagta
Long	5401	ttaaaaaaatctatatatcgctagtagtaaatgttaataaagtctctatttaaaacttgattttcatatgaaaaaaaaa

FIG. 4F

[illegible]

FIG. 5

21/51

EIF1AX & *EIF1AY*

EIF-1AX -207 ggcaacgagggcgccatttggctgccgcgagcg

EIF-1AX -130 tggacgcagcggtatctctgaagagctgggtgccagcctctcccgccac...gcctg.cctc.agcaccta.ttg...ccgc....t.
EIF-1AY -176 agttatgagagagctctgttagccagcctctctctg

EIF-1AX -86 ..t.g..tc..cc..cg.ag...ca.c.g.cgc.gtgcgcgtac..g.....a.....a.....a.g.cgcga.tc.c.gc.....
EIF-1AY -90 ccacctgctgcattcttagtgcgttcagtcggctcttagagtagtagtaaccgccagaaaggaggtcgcagaggtctcagaggtgtcatcacccgc

EIF-1AX 1 ..
EIF-1AY 1 ..

EIF-1AX 31 ..
EIF-1AY 31 ..

EIF-1AX 61 ..
EIF-1AY 61 ..

EIF-1AX 91 ..
EIF-1AY 91 ..

EIF-1AX 121 ..
EIF-1AY 121 ..

EIF-1AX 144 ..
EIF-1AY 144 ..

FIG. 6A

[illegible]

FIG. 6B

DEFRX & DEFRY

-1664	DIFFRY	gaaglgacatgttgccatgggccaattctgctggctcctttagt
-1620	DIFFRY	atacaaaaaataaagggttaccagtatgtcactacatgcagatttatggatgtacagaaaattggtgatcccaaatctcactgtgc
-1530	DIFFRY	atcaaaaataatcgatggaactttaagactaaagatttttagacccccaccaggcccgatgattgagaaatatctagaggggacccaaga
-1440	DIFFRY	atccatatatttaagtgcgcccccacaacaatgaccttaagcaggtagtcttgcatctgggaaccactgctacaggttactagtgggac
-1350	DIFFRY	aaccagltaggagcataagttgaacattttacagtttgcacctgtgatagtcttatcacctgtgatataaccagaaatccaattaaagt
-1260	DIFFRY	tgctatctctgttaactgttgcgaatttaggtgttaatttttgaagttcagaaaaagttagcaaaaacagaaaagaaatcaagta
-1170	DIFFRY	caactacataatgacaaaaaocgttatcacactgttatataacttcaaaactggagaataaaaggtgcgaataataacataataataat
-1080	DIFFRY	gctaagtgaataatataatcaaatgtagttgacctgaagaaaatcgactagtgaggatccctaacctgtgggccccccaggaatttactgt
-990	DIFFRY	tgaatggcttgagaatccactggaaaaagaccaagcatgtgtacctgaataatggaacttgtttattctccatatatttgcagtggtta
-900	DIFFRY	attccattataaaacctaatgaaaaaatgttttataagatgggtgtggaagactttctcgggctcagaggtgaaactgacctgtggtat
-810	DIFFRY	cagcagcatlctcagctgactgagugagtgtagtgattaacagagttgtgatgttagttaagaaaaacttagatttgccattgtagcttttc
-720	DIFFRY	taccaattagcagattgttttaactcactgaaattgtaaaagtgttagacgtggacttagtcatctactggcagcttatgaaattgtattcat
-630	DIFFRY	ttactcatgatgtaaaaatggttagttctccacttttaaggctctagtcttaglggctaaatagggtacttatttatcacagtatgataactg
-540	DIFFRY	ctgtalaaaaatcatgtctcaaatgtggaatagtagaagaggtgaaagaaatcatagtttgaggtagaatactgttctgctgggtcttaaa
-450	DIFFRY	aactgtggatttttggtgatccataaaataggltcagatacttccactggagggaacacagltttaagggataatagtgtactaatatag
-360	DIFFRY	aatgaggaagacacaccagatatttaggagggaatttagcgagcttgaaactaagagctgggttttgaaatagagctgggtcatcaagtgatttc
-270	DIFFRY	aagtaccagataagccactgagattttatttttaagcactgaagtcagatttttcccttttaaaagaaaggatttcatatgaaattctgc
-180	DIFFRY	tttttgcttgcagagagcttggagataaattcttggtggctgtlggagtatgtgtlggaggtattaaattttcacagtatataaaggca
-99	DIFFRX	c.tttctt..ag.ca.ctac.t...gc...c...tt.....cc.....g.....
-90	DIFFRY	gcaattgataggcctttcacagattcttctgataactcactataagagacaaaaaagagcaaaagttctgtgctgltgltcaagt

FIG. 7A

[illegible]

FIG. 7C

563 DFFRX T P V
 1687 DFFRY CTGAAACAAATAAGAGAAATTTGAGTTTGGTTGAAGCATCTCAAAATTTGAGTCAAATTCAGCGAAGTCCACACATATTTTATCGC
 571 L K Q I R F I C S L F G E A S Q N F S Q T Q R S P H I F Y R
 593 DFFRX C H V E M
 1777 DFFRY CATGATTAAATCAACAGCTTCAACAAATCAATGCTTTAGTTACTTTGGTAGCAGAAACCTTGCACCTACATGAATAGCATCAGATTG
 601 H D L I N Q L Q Q N H A L V T L V A E N L A T Y M N S I R L
 623 DFFRX R T G T G
 1867 DFFRY TATGCTGGAGATCAIGAAGACTAIGATCCACAAACAGTGAAGCTTGGAAAGTCGATACAGTCATGTTCAGAAAGCTTCAAGAACGACTAAAC
 631 Y A G D H E D Y D P Q T V R L G S R Y S H V Q E V Q E R L N
 653 DFFRX L T A A T
 1957 DFFRY TTCCCTTAGATTTTGTAGTGAAAGGATGGCCAACTCTGGCTCTGTGCTCCCTCAGGCAAAACAAATATGGAAGTCTTAGCAGAAATGCAGTT
 661 F L R F L V K D G Q L W I C A P Q A K Q I W K C L A E N A V
 683 DFFRX C T G A A
 2047 DFFRY TATCTTTGTGATCGTGAAGCCTGTTTAAAGTGGTATTCAGTTAATGGGGGATGAACCCAGACTTGGATCCCTGATATTAATAGGACTTC
 2071 Y L C D R E A C F K W Y S K L M G D E P D L D P D I N K D F
 713 DFFRX G T G G C
 2137 DFFRY TTGAAAGTANIGTACTTTCAGCTTGGATCCCTTTTAACTGAAATGGAATGAAATGCTTTGAAAGATTTTCAAGCTGTCAATGT
 721 F E S N V L Q L D P S L L T E N G M K C F E R F F K A V N C
 743 DFFRX G A G A T C
 2227 DFFRY CGAGAAAGGAAACTAATAGCAAAAGACATCCCTATATGATGGATTTGGAAATTAATGGACTAGACTACCTTTGGAGGGTGTGAT
 751 R E R K L I A K R R S Y M H D D L E L I G L D Y L W R V V I
 773 DFFRX N C C T Q
 2317 DFFRY CAGAGTAGTACCGAGATTCGCTAACAGAGCTTATAGATCTTCTTAAGAGATATACAAACCTTGGCCCAAGATTAAGCCCAATCAGGTG
 2341 Q S S D E I A N R A I D L L K E I Y T N L G P R L K A N Q
 781

FIG. 7D

FIG. 7E

1043 DFFRX DFFRY
 3127 S . . . T I . . . A . . . I . . . A . . . C . . . S . . . S . . . T . . . S . . .
 3151 GATAGAACAGCTGTAGAAAAATTAAGAGCTGTTTGTGTTGGACCAATGCAAAATAGTCCACCCCTTGACTCTTT
 1051 D R T A V E K L R A V C L D H A K L G E G K L S P P L D S L
 1073 A . . . G . . . T . . . C A . . . A . . . A . . . C . . .
 3217 TTTCTTTGGTCTCTCTGCCCTCCCAAGTTCTATACCTAACAGAGAGGTTAGTTTATGCTTGTGTTTAAATGCTGCTGGTGTGCTTAACTTCTTGGCAATGATGG
 1081 F F G P S A S Q V L Y L T E V V Y A L L M P A G V P L T T D G
 1103 A . . . G . . . T . . . C T C . . . A . . .
 3307 TCCCTCTGACTTTTCAAGTTTCACTTCTTCTTGAAGTGGTGGCTTACCTCTTGTACCTGAGTATGCTAATAAGAAATAACTTCTTGGCAAAATACA
 1111 S S D F Q V H F L K S G G L P L V L S M L I R N N F L P N T
 1133 C . . . C . . . C . . . G . . . T . . . C . . . T . . . T . . .
 3397 GATATGGAAACTCGAAGGGGTTGCTTATTTAAATGCTTCTTAAATAGCCAAACTGTGTGTAACTGGCTATGGCTATGGCCATGTTCGAGCT
 3421 D M E T R R G A Y L N A L K I A K L L L T A I G Y G H V R A
 1163 G . . . G . . . E . . . V . . . N . . . M C
 3487 GTAGCAGAAGCTTGTTCAGCCAGTTGTAGATGGTACAGACCCCAATACACAGATTAACCAAGTTACTCATGATCAAGCAGTGGTGTACAA
 3511 V A E A C Q P V V D G T D P I T Q I N Q V T H D Q A V V L Q
 1171 A A M V R C
 1193 AGTGCCTTCAGAGCAATTCCTAATCCCTCAATCCGAGTGCCTTACCTAGAAATGAGTCCATACTTCTGCTCAGGAAATATCTAATGAGGCT
 3577 S A L Q S I P N P S S E C V L R N E S I L L A Q E I S N E A
 1201 A A A G S Q
 1223 TCAAGTATATGCCCTGATATTTGTGTAATTAGGGCTATACAGAAAAATATCTTGGGCAATCAGCAATGTCGGCATAGGACTAGTTTATGC
 3667 S R Y M P D I C V I R A I Q K I I W A S A C G A L G L V F S
 3691 C E K A G P C
 1253 CCAATGAAGAATAACTAATAATTTATCAGATGACCCCAATGGAGCAATTAAGCTGGAGGTGAAGATGAACAAGTTTGTGTGAAGCA
 3757 P N E I T K I Y Q M T T N G S N K L F V E D E Q V C C E A
 3781 C G C C C
 1261 G G G G G

FIG. 7F

FIG. 7C

SUBSTITUTE SHEET (RULE 26)

FIG. 7J

FIG. 7K

FIG. 7L

35/51

CDYa & CDYb

CDYa	-281	ctgtggattta
CDYb	-328	gtaucaggcaggaagaagctttctgtactacaccagaggggttggggg.....
CDYa	-270	gctactctcacctgaggctactgagcaagttgtcatgcaccatgagacaaaagcccaagctgtgtccaccaggcagtaagtatggagaggtt
CDYb	-270g.....
CDYa	-180	caggcacatggcatagctgctatttcgcacaattttcactacaccagtggtgacaaaatagaagaggttcatccatcacacagaacacctggt
CDYb	-180C.....
CDYa	-90	gaagagctggaggcagaaaagaagtgtctatgtggagacgcaactgaacaaagggtggcacagcaactgttccaatcccggtgtcttctcctc
CDYb	-90
CDYa	1	M A S Q E F E V E A I V D K R Q D K N G N T Q Y L V R W K G
CDYb	1	ATGGCTTCCCAGGAGTTTGAGGTTGAAGCTATTGTTGACAAAGACAGGATAAAATGGGAATACACAGATATTGGTTCGGGTGGAAGGT
CDYa	31	Y D K Q D D T W F P E Q H L M N C E K C V H D F N R R Q T E
CDYb	91	TATGACAAACAGGATGACACTTGGGAACAGAGCAGCACCTCATGAACCTGTGAAAATGTGTACATGATTTTAATAGACGACAGACTGAA
CDYb	91A.....
CDYa	61	K Q K K L T W T T T S R I F S N N A R R R T S R S T K A N Y
CDYb	181	AAACAGAAAACAGTGCATGGACTACAAACAGTAGAATTTTTCAAACAATGCCAGAGAAGAAGAACTTCCAGATCTACAAAAGCAAACTAT
CDYb	181
CDYa	91	S K N S P K T P V T D K H R S K N R K L F A A S K N V R R
CDYb	271	TCTAAGAACTCTCCATAAACGCCAGTGCATGATTAACACACAGGTCCAAAACCCGCAAGTTATTGCTGCCAGCAAGAACGTTAGGAGA
CDYb	271T.....
CDYa	121	K A A S I L S D T K N M E I I N S T I E T L A P D S P F D H
CDYb	361	AAGCAGCTTCAATCTCTCCGACACAAAGATATGGAGATAATAATTCAACTATTGAGACCCCTTGACACCTGACAGCCCTTTTGACCCAC
CDYb	361C.....
CDYa	151	K - T V S G F Q K L E K L N P I A A D Q Q D T V V F K V T E
CDYb	451	AAA---ACTGTAGTGGCTTTCAGAACTTTCAGAAAACCTTATTCAGACAGATCAGCAGGACACGGTGGTCTTCAAGGTGACAGAA
CDYb	451G.....
CDYa	180	G K L L R D P L S R P G A E Q T G I Q N K T Q I H P L M S Q
CDYb	538	GGGAACCTCCCTCCGGACCCCTTGTCTCAGTCCCTGTCAGAACAGACTGGATACAGAACAAAGACTCAGATACACCCACTAATGTGCGCAG
CDYb	541A.....
CDYa	210	M S G S V T A S M A T G S A T R K G I V V L I D P L A A N G
CDYb	628	ATGTCTGGCTCAGTTACTGCTTCTATGCCACAGGTTTCAGCTACCCGAAAGGGTATAGTGGTATTANTAGACCCATTAGACCCCAATGGG
CDYb	631
CDYa	211

FIG. 8A

FIG. 8B

37/51

BPY1

gagaggggtatcacagggaggccagcctggaggttagtcgaccgttgcgagacgttgagctgaggcag
 -72
 1 ATGAGTCCAAAGCCGAGAGCCTCCGGACCTCCGGCCCAAGCCAAAGGACACAGCAAGAGGAAGTCTCTCTCAGCCGAGCCCCAGTGGC
 1 M S P K P R A S G P P A K A K E T G K R K S S Q P S P S G
 91 CCGAAGAGAAGACTACCAAGGTGCGCCGAGCAAGGAGAGCAGTTCTGTGGAGGGAGACGCGGGAAGAAAGGGGCTCCGACAAAGATGCGG
 31 P K K T T K V A E K G E A V R G G R G K K G A A T K M A
 181 GCCGTGACGGCACCTTGAGGGGAGCGCGGCCAGCCCGGCCAGCCAGCCAGCCAGGAGCTCCCTCAGCAGCAGGCTGCCG
 61 A V T A P E A E S G P A A P G P S D Q P S Q E L P Q H E L P
 271 CCGGAGGAGCCAGTGAGCGAGGGGACCCAGCACGACCCCTTGAGTCAAGGAGAGCGAGCTGGAGGAAACCACTGAGTAAGGGCGGCCATCT
 91 P E P V S E G T Q H D P L S Q E S E L E P L S K G R P S
 361 ACTCCCTATCTCCCTGAGcagcancataagtttagggccagctgccagacctcagagatctcaccagcagggtgcttcccatgttgatga
 121 T P L S P * 125
 451 caataaatgaatgtgttgcaaaaaaaaaa 480

94

FIG. 9

38/51

BPY2

-332 aatatctcaggaccaggaccatgtgatatggggcccaaacacctggatgatgttactcttctg
 -270 cctaggtcatgcgtaaaagagggaattagggccatattgcttgcccgagtcctgactctcctgcttgccagagccacaga
 -180 agtggcttggtgacataaactcttgaggctgtcacatcaccaagattatatgttactgactggaccagcataaaagctgacacttctgacta
 -90 tggccagccttcaataatactacactgtataattggctcaacaccagggtgatatgtgtccatttactgagaccagataaaaaagccta

 1 ATGATGACGGCTTGTCCTCCAGAGCCAGGACACGTCAGGACAGGATCATTTACTCTCATTCCTTGCCTCCAGATTTTCACAGGTGCTGCTTACA
 1 I M M T L V P R A R T R A G Q D H Y S H P C P R F S Q V L L T
 91 GAGGGCATCATGACATAITGCTTGACAAAGAACCTTAAGTGATGTAAATTTCTCCATAGGTTCCTTAAATAATGGGAATGTGAGAAATACC
 31 E G I M T Y C L T K N L S D V N I L H R L L K N G N V R N T
 181 TTGCTTCAGTCCAAAGTGGGCTTGCTGACATATTTATGTGAACCTGTACCCGGGTGAAGTGACTCTTCTGACTAGGCCCCAGCATACAAATG
 61 L L Q S K V G L L T Y Y V K L Y P G E V T L L T R P S I Q M
 271 AGATTATGCTGTATCAGTGGCTCAGTCTCGAAGCCAGATCACAGAAAGTAATTgtgccatatgtggaacaagcagctaagcaatagataa
 91 R I C C I T G S V S K P R S Q K * 106
 361 catccatcgtggctctgccttcaaaagggaattttacatatgtcactgggaccatcacccagatgatgtcctgccccactaaaaagaattgt
 451 gacataacgctgactgcaaaaactgggtaattgcaactctcctcttattctggagcttgccaaacaaggattatcacatatgtcgaggag
 541 tccagcaccaggtaaaatttctcatataccagcttcagataccatgcaatgatatacaactatcatacctggacccaaaggagagagat
 631 atttggattctcattgccaattcttatggccacaagcaagtaattggtctcatagtggtataaagttcacacagttattatgacactccca
 721 gcgtatcatagaaaaatgtgagtagtagtaacaatgagtggtataacagggaacagcaaaccaatgcttattgtgatttggatttcacacccagc
 811 tgacgcgactatcatctctcacaaagacagaaacctgcgaataaaagtactactaaatctcaccaaaaaaa 880

FIG. 10

[illegible]

FIG. 11

40/51

PTPRY

```

-182      gaagaggagcacaccacacagagaacacacatctgtcagtggttcactgtctcaaccttatctgcacagtcgaggtcagtcctgagagag
-180      ctctgagagacccaggatgaaggatgcagtgaggtcaagagcccaacctcttctactgacacccacctctaaggactcagaagagac
-90
1  ATGAAATATAATGGGCTCAACAATCCCAAGAGAACCACACTCAAGGACAATGGGAGCCACTGGGCTTGGCTTCTACTTCCCTGGAACAA
91  M N K M G L N N P K K N H S R T M G A T G L G F L L P W K Q
31  GACAATTGAAATGGCACATGACTGCCAGGGATGCAATATTTATACTTCTCTGAGACTACGGGGAGCATGTGTCTGAACATTTCCCTGAAC
61  D N L N G T D C Q G C N I L Y F S E T T G S M C S E L S L N
181  AGAGGTCCTTGAGGCCAGAGGAGGATCTTAAGACTCATTTCTGAGAGATATGGGAGGTGTGGCTGTATCTCACTTCCACTTTCGT
61  R G L E A R R K K D L K D S F L W R Y G K V G C I S L P L R
271  GACATGACCGCTGGATTAACCCACCCCAATTTTCAGAGAATTTTCCAAAGGCTACCAAGGGTGGACGGAGCTGATGCACATGAGCCTG
91  E M T A W I N P Q I S E I F Q G Y H Q R V H G A D A L S L
361  CAACCAACTCTCTGAGAGCAGGTATCTTCAAGTCCCTCGGACAGAGCTTCTTCTCAGGACACTCGAGAGAGCCGTGGTTTCAGGG
121  Q T N S L R S R L S S Q C L G Q S F L L R T L E R A V V S G
451  CACTTGGGACATCTGTGGCCACGTTCTATTAAGAGAGACTAAGCCTACTTCTCAGGACCCGCCAAGAGTGCCCGGCTTTGGGACA
151  H L G T S V A T F M K K T K P T S S Q D P P K S G R G F G T
541  CCTGCGGTGCGGTCCACCATGAGGATATAAACCCTTCTTCTTCTGACATGTCCAGGAGTGGCCGTGTCTACAAAGTCACCTGGTGCTACG
181  P A V G S T M R I K P P S L L D M S R S G R C Y K S P G A T
631  ACCAGGTTGAGATATAAGACGTCCTCCACAGACCCCTCCAGGAGAGTACATGGCATTTGACACATCTGGCGGCCAAGTGAGGAAAGACAC
211  T R V R I K T S P Q D P P R R V H G I E T S G G Q V R K R H
721  CCTGCTGCAGCACCCAGAACTGAGggggggcactgccctgggccttacttccagccctggcctcccaattctgaccttacaaaaagtgtc
241  P V C S T Q N * 247
811  ccttgagtgaggcagtgaccacgcattgtcacagctaccacaaagtgtgttctgagatgatctgggcttgttcttggcagagattctggta
901  cagagaaaggagagcgcttgagtggaaccacgatgggctgagggccaggaggagacatcacacccctcccaaacacttttttcattgctta
991  ataaatcatttttcttagagaactaaagtgtgaaacaatatagaaacattttttaagttaggcataaaaaaa 1066

```

FIG. 12

41/51

TTY1

tgctgtcagagctgtcagcctgcttaagcagagtaaaatgggtacaggcagtgtagcagcctgggtagcgagaaaaaaggctgcctgtgaaatc
ccactgtgggaccataaagtggggacctcagggcccttccatggcatctccatggccatgtcatgtctgggagaaggagcggtttcaagaatg
tgagctgacgcgtggaactgtctcatctgactccagtcctcaaaaggagctatgtgcaagaatcggtgaagtgtgagaccccatccacc
cctcaagaattgtatccccaccctgtctgaccttactgtctgacctatctgtccaaggatgaaacccaggacacaaaggaggagtaa
ccctcatgatgtgaagcacgtgtcacctgtgaataaaccctgaggatcatgagactatctgtggatttcacagagaagacagacgagaa
gacaccgtgacacttctccacggagggtctcttccaccagaatgcagatgcttcttgcaaggactatcctgtgaatccacacagagaa
gacagggtgggttccaaagccgggtgcacctccagggaattctccttctctaccaggctccaggccttctgccatgatcatgagactattt
gtggatttcacagagaagataggtgaaggtacagcatgggcatccacccctcaccagaggggtatccccaccctatctgacctattacc
ttattgtgttcaaaagtctctatccagactgaaatcccaagacaaatggagaagtccccctgagtgtgaagcaccaactcctctggg
aatcaaatcgagggtaaattlaatagggccggtagagatgaatgatgtgtctctctctctgggctgaaagacaaattaaacactggta
tattctgtttaaaaaaaaaa

FIG. 13

$$T^TY2$$

FIG. 14

43/51

Human CDYL (CDY Like)

ggagaggacctatttctacctaaggacatbccggaaggcaatgggttcaaacaaatatacct
gaagagactcatctcggggaactaagcaggtgtaatacagagaacacagagccccggaagaat
tttATGGCATTTTCAGGCAAGCCACAGGCCAGCCTGGGAAAAGCAGGAAGAAAACCTGGCAAT
ACGAGGGCCCAACCCAAAAGTTATTCTGAAGAGAAAACACGTGTCAGCACCCAGATGGGCCCTTC
AGACCCAGCATCTCCGCGAGCAGTGAGCAAAAGCGGGCACAGCAGCCTCCCCGGTTTACAGGTT
GAAAGGATTGTTGACAAAAGGAAAATAAAGGGAAGACAGAGATAATTGGTTTCGGTGGAAAG
GCTATGACAGCGAGGACGACACTTGGGAGCCGGAACAGCACCTCGTGAACCTGTGAGGAATACAT
CCACGACTTCAACAGACGCCACACGGAGAAAGCAGAGGAGAGCACATTGACCAGAAACAACAGG
ACCTCTCCCAACAATGCTAGGAAACAATACTCCAGATCCACCACAGCAACTTTCTAAGACCT
CTCCTAAGGCACCTCGTGATTGGGAAAGACCACGAATCCAAAACAAGCCAGCTGTTGCTGCCAG
CCAGAAAGTTCAGGAAGAACACAGCTCCATCTCTCCAGCCGGAAGAACATGGACCTAGCGAAG
TCAGGTAACAAGATCCTCGTGCCATAAAGCCCCGTTAAGAGCAGGACCGCAGTGGACGGCTTTC
AGAGCGAGAGCCCCTGAGAAACTGGACCCCGTCGAGCAGGTCAGGAGCACACAGTGGCACCCGA
AGTGGCAGCGGAAAAGCCGGTCGGAGCTTTATTGGGCCCCCGTCCCGAGAGGGCCAGGATGGGG
AGCAGGGCCAGGATACACCCACTAGTGCCCTCAGGTGCCCGGCCCTGTGACTGCAGCCATGGCCA

FIG. 15A

44/51

CAGGCTTAGCTGTTAACGGGAAAGGTACATCTCCGTTTCATGGATGCATTAAACAGCCAATGGGAC
AACCAACATACAGACATCTGTTACAGAGTGAAGTCCAGCAAAAGGAAATTTATTGACGACAGA
AGAGACCAGCCTTTTGACAAAGCGATTGCGTTTCAGCGTGAGCAAAACAGAAAGTGCCCTACAGAT
ACAGAGATATTGTGGTCAGGAAGCAGGATGGCTTCACCCACATCTTGTATCCACAAAGTCCCTC
AGAGAACTAACTCACTAAATCCAGAGGTAATGAGAGAAAGTCCAGAGTGCTCTGAGCACGGCCGCT
GCCGATGACAGCAAGCTGGTACTGCTCAGCGCCGTTGGCAGCGTC'TTCTGTGTGGACTTGACT
TTATTATTATACGACGCTGACAGATGACAGGAAAGAGAAAGCACTAAATGGCAGAAAGC
TATCAGAAACTTCGTGAATAC'TTCAATTCAATTAAAGAGCCCATTTATTGTAGCAGTCAATGGC
CCAGCCATTGGTCTAGGAGCATCTATATTGCCCTCTTTGCGATGTGGTTTGGGCTAATGAAAAGG
CTTGGTTTCAACACCCATATACCACCTTCGGACAGAGTCCAGATGGCTGTTCTACCGTTATGTT
TCCCAAGATAATGGGAGGAGCATCTGCAACAGAGATGCTGCTCAGTGGACGGAAGCTGACACGG
CAGGAGCGGTGGCAAGGGCCCTGGTCTCCAGGTGTTTGGCCCGGACGTTCACTCAGGAAG
TGATGGTTCGCATTAAAGGAGCTTGCCCTCGTGCAATCCAGTTGTGCTTGAGGAATCCAAAGCCCT
CGTGCGCTGCAACATGAAGATGGAGCTGGAGCAGGCCAACGAGAGGAGTGTGAGGTGCTGAAG
AAAATCTGGGGCTCGGCCCCAGGGGATGGACTCCATGTTAAAGTACTTGCAGAGGAAGATCGATG

FIG. 15B

45/51

AGTTCTGAgTgtcgggctgccactggtgacaccgggatcgggctgagcaggagaacatcacccg
gctccagttcccctgatccattctcacagcctgaacaagctcacccgtagcttacgcttgga
gcaggactgggaacatccacgctatttattatcgaggagttaaagtactgtaactttaaaat
aaataactacaagcttcttgtcvaacgtcattatttatacttatatacacgcaggtgtaa
aagtataaaggtagcactagactgctcttagaagctctaatttttgtttcttggctagtac
tgtataaaaaacagaaattgtgttttatgtgttttggtgacagaaaaagtctggaataatgtttg
tttccctcatcttccctcttagaacacagaaatctaaggggtgttagccagcctcgccctccct
gcccacgtagagacacagagtgtgaggcgttggttttctccaagaaggtaacagatacc
tcagattcgggaactcaaaatcaaaagacttagcttctaggataaaatacttctgatgaaaaat
ccgtgaggagcataccccaaaccagacatatgtcttaggattcatgctgagatatcaattgggtt
tccccttctttaaatalacgtccagttcttaccagttaacatgaagaaccactgtctcttag
aagaaagcttggtttgcagtattagtgaatcactgaatagcttaagtatgactatctaagttat
aagttagttcttagtgggttttaaatagtttctgaccttctgaaaaataactacataagt
cttctgttgctgggtgagaaatactactttatagacagttttgggtttctgtttgcagatatg
attgatgtatttcaccaaaaaataaataatttttatgtttataaagtgtaattttttaggttcactt
agaatatattttatttaataagttaaaaattcttttggcacactatttaaatgcaaaaaactcctt

c

FIG. 15C

Mouse *Cdyl* (CDY like)

cttbtgagtggtttagcatcccacttgttcctbtgaggacatctgttctactaagagcactcacc
tgagatgctcaagggtccagaagaacacttctcggtgacaaagcaggtggtgaccagagaacag
aggcccccaaaattttatggcattcaaggcaagcacagccaaacccggagggaaagcaagagtc
cagcctggaaatacatagcccacccgaaggttatctctgaaggaaaacaattggcattagGCAATA
GCCAGCCTAATTACAGGAAGCCAGCTCTGCACACTTCCAGAGAAAGCTGAACAACCTACTGATG
ATAACACCTGCCAGCAAAATAATGTGGTTCTTGCAACAGTCTCAGAACCCGATCAAGCGTCCCCCTG
CAATTCAAGACGCGGAGACTCAGGTGGAAAGTATCGTTGACAAAAGGAAAACAAGAAAGGAAGA
CAGAAATATCTGGTGGGTGGAAAGGCTATGACAGTGAGGATGACACGTGGGAGCCTGAGCAGCAC
TGGTGAACCTGTAGGAATACATCCATGACTTCAACCGCGCCACACGAGAGGCAAAAGGAAGGTA
GCCTGGCTCGTGCCAGCAGAGCCTCCCCCAGCAACGCCCGGAAAGCAGATTTCAGGTCCACCCACA
GCACTCTCCAAGACCAACTCCAAGCACTTGTGGTAGGCAAAAGATCATGAGTCCAAAAGCAGCC
AGCTGTTGGCTGCCAGCCAGAAAGTTCAGGAAAACCCAGCCCCATCTCTTGCAAAACCGCAAGAAC
TGGACCTCGCCCAAGTCAGGGATCAAAATTCCTGTCCTAAGAGCCCCCGTTAAGGGCAGGACCTCGG
TTGATGGCTTTCAGGGGGAGAGCCCCCGAGAAAGCTGGACCTGTGGATCAGGGTCCCGAGGACACTG
TAGCCCCAGAGGTGACTGCAGAGAAGCCCCACTTGGGGCTTTGCTGGGCCCTGGTGGGAGCGAGCCA

FIG. 16A

GGATGGGAGCAGGCCCCGAATACATCCACTAGTCCCTCAGGTTTCTGGCCCCGTGACTGCTGCCA
TGGCCACAGGCTTAGCTGTTAATGGAATAAGGTACATCTCCATTGATGGATGCGTAGCAGCCAACG
GAACAGTCAACCATACAGACATCCGTAACAGGAGTGACAGCCGGGAAAAGGAATTTATTGACGACA
GAAGAGACCAACCTTTTGACAAAGCGGTTGCGTTTCAGTGTGAGCGACACAGAGAGTGCCCTACAGAT
ACAGAGATATTGTCGTCAGGAAGCAAGATGGCTTCACCCACATCTTGTTATCCACAAAATCGTCAG
AGAAATACTCACTAAACCCAGAGGTGATGAAGAAGTRCAGAGCGCCCTGAGCACAGCTGCAGCCG
ACGACAGCAAGCTGGTTCTGCTCAGCGCCGTGGCAGCGTCTTCTGCTGTGGTCTGGACTTTATTT
ATTTTATTTCGGCGCCTCACAGATGACCGAAAGAGAGAAAGCACTAAAATGGCAGACGCTATCAGAA
ACTTCGTGAATACTTTCATTTCAGTTTAAGAAAGCCTATTATTGTAGCTGTTAATGGCCCAAGCATTG
GACTAGGAGCATCCATAATTGCCCTCTTTGTGATGTGGTTTGGGCTAACGAAAAGGCTTGGTTTCAAA
CACCCTATACCACTTCGGACACAGATCCAGATGGCTGCTACCGTTATGTTTCCCAAGATTATGG
GAGGAGCATCTGCGAATGAATACTGCTGTTTCAGTGGCGGAAGTTGACGGCACAGGAGGCTGTGGCA
AGGCTCTGGTCTCCAGGTGTTTGGCCAGGAACCTTCACACAGGAAGTCATGGTTCGAAATCAAGG
AGCTGGCTTCATGTAAACCCAGTTGTCCTGGAGGAATCCAAAGCCCTGGTGGCTGCAATATGAAGA

FIG. 16B

TGGAGCTAGAGCAGGCCAATGAGAGAGAAATGTGAAGTGTGAAGAAGATCTGGGGCTCCGCCCAGG
GCATGGACTCCATGTTAAAGTACTTACAGAGGAAATCGATGAGTTCTGATgggcaggctgagcag
gacatcgggtggctcccacttgctacgtcgtcctgcagtggtcgtgcttgaggcagaactggaaa
catccgagctatttattgcccggagtttttaagtaactttaaaataatacaaaagcttct
ttgtctaagcgtcttattttatactcatgtatacacaaagtataaaaaatgtaattgagcactaggc
tgctcttggaagctctaatttcttgtaagctagttgtggattttgttttggtttttggttttaaa
aggaaattatgttttcattttgggtgacagaagagtttgaaataatgtttgttttactcttttttt
tttccttaaatctagatcacagaccctcaaaattactagccagccttctccccctctactga
aacatgtagaaataacttaaacatgttctgcctctaggggggagggtgtgagtcacctcaat
gctgaaaaacagttctgatcaaaccttaagaccaacctggtaaaaaaagcatcactgatggaaaaatcc
caccacggggcggtgttctgtgaaatgccgcgctctacctttcttactgtcccatctt
accagccaccgtgaagagcccagtgcttgaggaaagcagggtggtccagtgctgtgagtcactc
cgtagctcgagtggtacttgctlaagttatgaattagcatttagtgggtttaaatagtttttctgacc
ctttttgaaaaataactacataagtaactccttggtggctgggtgagaaatactacttttgcatagttt
tgtttgctctatctgcagatatgatgtgtgtattacaccaaagtaatttttatgtttataaagtgt
aattttttaggttcacttagaataataattttatttaatttaaaattctcttggcacactattaaatac
gtaaacctcctttc

FIG. 16C

VCP2r (VCP with 2 repeats)

gttgagacgttgagctgcggaagATGAGTCCAAGCCGAGAGCCCTCGGGACCTCCGGCCCAAGGCCAC
GGAGGCAGGAAAGAGGAAGTCCTCTCTCAGCCGAGCCCCAGTGACCCGGAAGAAAGACTACCAAGGT
GGCCGAGAAGGGAAGCAGTTTCGTAGAGGGAGACGCGGGAAGAAAGGGCTGCCGACAAAGATGGCGGC
CGTGACGGCACCTGAGGCGGAGAGCGGGCCAGCGGCACCCGGCCCCAGCGACCCAGCCAGCCAGGAGCT
CCCTCAGCACGAGCTGCCCGCGGAGGCCAGTGAGCGAGGGACCCAGCACGACCCCGAGTCAGGA
GGCCGAGCTGGAGGAACCACTGAGTCAGGAGCGAGGTGGAAGAACCACTGACTGTGTGGATGGCCAG
CTTTTCCCCCTGTC'TCCGAGAGCAGCGACTAAgttcaggccccagcccgacacctcagagatctcaccag
cggggtgcttgccattctgaagataataaaatgaatgtgttgcaaattgaaaaaa

FIG. 17A

50/51

VCP8r (VCP with 8 repeats)

cggaagATGAGTCCAAAGCCGAGAGCCTCGGGACCTCCGGCCCAAGGCCACGGAGGCAGGAAAGAGGAAG
TCCTCCTCTCAGCCGAGCCCCAGTGACCCGAAGAAGAACTACCAAGGTGGCCCAAGAAAGGAAAGCA
GTTCTAGAGGGAGACGCGGAAGAAAGGGCTGCGACAAAGATGGCGGCCGTGACGGCACCTGAGGCG
GAGAGCGGCCAGCGGCACCCGGCCCCAGCCAGCCAGCCAGGAGCTCCCTCAGCACGAGCTGCCG
CCGAGGAGCCAGTGAGCGAGGGACCCAGCACGCCCTGAGTCAGGAGGCCGAGCTGGAGGAACCA
CTGAGTCAGGAGCGAGGTGGAAGAACCACTGAGTCAGGAGAGCCAGGTGGAGGAACCACTGAGTCAG
GAGAGCGAGGTGGAGGAACCGCTGAGTCAGGAGAGCCAGGTGGAAGAACCACTGAGTCAGGAGAGCGAG
GTGGAGGAACCACTGAGTCAGGAGAGCCAGGTGGAGGAACCACTGAGTCAGGAGAGCGAGATGGAAGAA
CTACCGAGTGTAGACGGCCAGCTACTCCCCCTATCTCCGAGAGCAGCGACTAAGttcaggccagccg
ccagacctcagagatctcaccagcgggtgtcttgccattctgaagataataaaatgtaattgtgtgcaaa
ttgaaaaaa

FIG. 17B

51/51

VCP10r (VCP with 10 repeats)

cgttcgagacgttgagctgcggaagATGAGTCCAAAGCCGAGAGCCCTCGGGACCTCCGGCCCAAGGCCA
CGGAGGCAGGAAAGAGGAAGTCCTCTCTCAGCCGAGCCCCAGTGACCCGGAAGAAAGACTACCAAGG
TGGCCAAGAGGAAAGCAGTTTCGTAGAGGGAGACGCGGGAAGAAAGGGCTGCCGACAAAGATGGCGG
CCGTGACGGCACCTGAGCGGAGAGCGGGCCAGCGGCACCCGGCCCCAGCGACCCAGCCAGCCAGGAGC
TCCCTCAGCACGAGCTGCCCGGAGGAGCCAGTGAGCGAGGGACCCAGCACGACCCCTGAGTCAGG
AGGCCGAGCTGGAGGAACCACTGAGTCAGGAGAGCGAGGTGGAAGAACCACTGAGTCAGGAGAGCCAGG
TGGAGGAACCACTGAGTCAGGAGAGCGAGGTGGAAGAACCACTGAGTCAGGAGAGCCAGGTGGAGGAAC
CACTGAGTCAGGAGAGCGAGGTGGAGGAACCACTGAGTCAGGAGAGCCAGGTGGAGGAACCACTGAGTC
AGGAGAGCGAGATGGAAGAACCACTGAGTCAGGAGAGCCAGGTGGAGGAACCACTGAGTCAGGAGAGCG
AGATGGAAGAACTACCGAGTGTGTAGACGGCCAAGTACTCCCCCTATCTCCGAGAGCAGCGACTAAGttc
aggcccgcccgagacctcagagatctcaccagcgggtgcttgcattctgaagataataaaatgaa
tgtgttgcaaattgaaaaaa

FIG. 17C

INTERNATIONAL SEARCH REPORT

International Application No

US 98/07115

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C12N15/11

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 92 00375 A (IMP CANCER RES TECH) 9 January 1992 see the whole document ---	1
X	ZHANG J. ET AL.: "Molecular isolation and characterization of an expressed gene from the human Y chromosome" HUMAN MOLECULAR GENETICS, vol. 1, no. 9, December 1992, pages 717-726, XP002080218 see the whole document ---	1,2
-/-		

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *Z* document member of the same patent family

Date of the actual completion of the international search

12 October 1998

Date of mailing of the international search report

17.12.98

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Kania, T

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/07115

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MA K. ET AL.: "A Y chromosome gene family with RNA-binding protein homology: candidates for the azoospermia factor AZF controlling human spermatogenesis" CELL, vol. 75, no. 7, 31 December 1993, pages 1287-1295, XP002017338 cited in the application see the whole document ---	1,2
X	WO 95 11300 A (MEDICAL RES COUNCIL ;CHANDLEY ANN CHESTER (GB); KUN MA (GB); SHARK) 27 April 1995 see the whole document ---	1,2
A	WO 97 10267 A (PROMEGA CORP ;KENT MARIJO G (US); AGULNIK ALEXANDER I (US)) 20 March 1997 see the whole document ---	1-4,8
A	PAGE D. ET AL.: "The sex-determining region of the human Y chromosome encodes a finger protein" CELL, vol. 51, no. 6, 24 December 1987, pages 1091-1104, XP002080219 cited in the application see the whole document ---	1-4,8
A	WO 96 41007 A (PROMEGA CORP) 19 December 1996 see the whole document ---	1-4,8
A	FOOTE S. ET AL.: "The human Y chromosome: overlapping DNA clones spanning the euchromatic region" SCIENCE, vol. 258, 2 October 1992, pages 60-66, XP002080220 see the whole document ---	1-4,8
P,X	LAHN B. AND PAGE D.: "Functional coherence of the human Y chromosome" SCIENCE, vol. 278, 24 October 1997, pages 675-680, XP002080221 see the whole document -----	1-4,8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 98/07115

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see continuation-sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-4,8 partially (subject 1. on continuation-sheet)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-4,8 partially

Isolated DNA which occurs on the non-recombining region of the human Y chromosome or the complement thereof, being testis-specific and optionally occurring in multiple copies on the Y chromosome.

Said DNA being the CDY gene, a characteristic portion, a probe or complement thereof, a DNA which hybridizes thereto under stringent conditions.

Said DNA having the SEQ ID NO:37,38 and coding for the amino acid of SEQ ID NO:39,40.

2. Claims: 1-4,8 partially

idem for BPY 1, SEQ ID NO:41,42

3. Claims: 1-4,8 partially

idem for BPY 2, SEQ ID NO:43,44

4. Claims: 1-4,8 partially

idem for XKRY, SEQ ID NO:45,46

5. Claims: 1-4,8 partially

idem for PTPRY, SEQ ID NO:47,48

6. Claims: 1-4,8 partially

idem for TTY 1, SEQ ID NO:49

7. Claims: 1-4,8 partially

idem for TTY 2, SEQ ID NO:50

8. Claims: 5-7,9 partially

Isolated DNA which occurs on the non-recombining region of the human Y chromosome or the complement thereof, not being testis-specific and having a homolog on the human X chromosome.

Said DNA being the DBY gene; a characteristic portion, a probe or complement thereof, a DNA which hybridizes thereto under stringent conditions.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Said DNA having the SEQ ID NO:17 and coding for the amino acid of SEQ ID NO:18.

9. Claims: 5-7,9 partially

idem for TPRY, SEQ ID NO:19,20,21,22,23,24

10. Claims: 5-7,9 partially

idem for TB4Y, SEQ ID NO:26,28

11. Claims: 5-7,9 partially

idem for EIF1AY, SEQ ID NO:30,32

12. Claims: 5-7,9 partially

idem for DFFRY, SEQ ID NO:34,36

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/07115

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9200375 A	09-01-92	AU 670229 B AU 8093191 A CA 2085102 A EP 0536213 A	11-07-96 23-01-92 29-12-91 14-04-93
WO 9511300 A	27-04-95	AU 7947794 A	08-05-95
WO 9710267 A	20-03-97	AU 7156896 A CA 2238694 A EP 0859790 A	01-04-97 20-03-97 26-08-98
WO 9641007 A	19-12-96	US 5783390 A US 5776682 A AU 6159296 A CA 2221521 A EP 0832288 A US 5840549 A	21-07-98 07-07-98 30-12-96 19-12-96 01-04-98 24-11-98